

**SUR L'ESTIMATION PONCTUELLE POUR DES
MODÈLES DE MÉLANGE**

par

Fanny Rancourt

Mémoire présenté au Département de mathématiques
en vue de l'obtention du grade de maître ès sciences (M.Sc.)

FACULTÉ DES SCIENCES
UNIVERSITÉ DE SHERBROOKE

Sherbrooke, Québec, Canada, 7 novembre 2019

Le 7 novembre 2019

*le jury a accepté le mémoire de madame Fanny Rancourt
dans sa version finale.*

Membres du jury

Professeur Éric Marchand
Directeur de recherche
Département de mathématiques

Professeur William E. Strawderman
Évaluateur interne
Département de mathématiques
et Rutgers University

Professeur Félix Camirand Lemyre
Président-rapporteur
Département de mathématiques

Sommaire

Dans ce mémoire, on s'intéresse à l'estimation ponctuelle pour des modèles de mélange, dénombrables ou non, sur les réels positifs. Premièrement, en généralisant les résultats obtenus par Kubokawa, Marchand et Strawderman [23], on obtient une classe d'estimateurs lisses dominant un estimateur sans biais sous la perte quadratique. On compte notamment les chi-deux et Fisher décentrées, la loi du coefficient de détermination expérimental R^2 en présence d'un modèle normal, ainsi que les modèles à seuil aléatoire tel que présenté par Pigeon et Denuit [40] parmi les applications. De plus, ces résultats s'appliquent en présence d'une contrainte inférieure de l'espace paramétrique. Enfin, on effectue une analyse plus approfondie de l'estimation du paramètre de décentralité d'une loi chi-deux. En effet, on compare les risques fréquentistes de l'estimateur de vraisemblance maximale tel que présenté par Saxena et Alam [46] à la classe d'estimateurs précédemment obtenue puis on conclue avec les inférences bayésienne et prédictive avec un *a priori* gamma.

Mots-clés: estimation ponctuelle, espace paramétrique restreint, modèle de mélange, comparaison de risques, paramètre de décentralité, chi-deux décentrée.

Remerciements

Je tiens premièrement à remercier mon directeur de recherche, M. Éric Marchand, pour sa disponibilité, son soutien et tous les échanges enrichissants, académiques et personnels, dont ma maîtrise a été ponctuée. Ces dernières années m'ont permis de grandir sur plusieurs plans de ma vie, notamment grâce à ses judicieux conseils.

Je veux également remercier mes collègues du Département de mathématiques pour leur soutien, pour les discussions mathématiques et statistiques ainsi que pour les moments plus ludiques.

Je tiens aussi à remercier pour leur soutien financier l'Institut des sciences mathématiques, le Département de mathématiques ainsi que le fond de recherche de mon directeur.

Enfin, je suis extrêmement reconnaissante du soutien indéfectible de mon conjoint, de ma famille et de mes amis.

Table des matières

Sommaire	ii
Remerciements	iii
Table des matières	iv
Liste des figures	vi
Abréviations et notations	viii
Introduction	1
1 Notions préliminaires et premiers résultats	4
1.1 Mélanges de lois de probabilité	4
1.1.1 Mélanges indénombrables	5
1.1.2 Mélanges dénombrables	7
1.2 Ordonnancement stochastique	13
1.2.1 Croissance stochastique	14
1.2.2 Rapport de vraisemblance monotone	16
1.3 Théorie de la décision statistique	19
1.3.1 Équivariance	21
1.3.2 Règles de Bayes et critère minimax	23
1.4 Méthode de Kubokawa	30
1.4.1 Pour les lois continues	31
1.4.2 Pour les lois discrètes	38

TABLE DES MATIÈRES

2	Estimation pour des modèles de mélange	47
2.1	Pour un espace paramétrique non borné	48
2.2	Pour un espace paramétrique borné inférieurement	62
3	Sur l'estimation du paramètre de décentralité d'une loi chi-deux	67
3.1	Sur les propriétés de l'EVM	68
3.2	Comparaisons de risques et résultats de dominance	72
3.3	Inférences bayésienne et prédictive	76
	Conclusion	86

Liste des figures

1.1	Fonction ψ_U et risques d'une normale de moyenne θ et d'écart-type 1 lorsque $\theta \geq 0$ pour la perte valeur absolue.	35
1.2	Risques pour la perte valeur absolue dans le cas d'une exponentielle de position de paramètre $\theta > 0$	37
1.3	Risques et rapports des risques pour la perte quadratique dans le cas d'une loi de Poisson de moyenne $\theta \geq 1$	42
1.4	Risques pour la perte quadratique dans le cas d'une géométrique de moyenne $\theta \geq 1$	46
2.1	Rapport des risques $\frac{R(\delta_{\psi^*}, \theta)}{R(\delta_0, \theta)}$ pour un mélange de lois inverse-gamma et gamma d'échelle $\theta \geq 0$	57
2.2	Risques $\frac{R(\delta_{\psi^*}, \theta)}{R(\delta_0, \theta)}$ d'une loi Fisher décentrée de degrés de liberté 7 et 5 puis de paramètre de décentralité $\theta \geq 0$	61
2.3	Rapport des risques $\frac{R(\delta_{\psi^*}, \theta)}{R(\delta_0, \theta)}$ pour l'estimation de $\frac{\rho^2}{1-\rho^2}$ pour un échantillon de taille 8, un problème à 4 variables et $\theta \geq 0$	62
2.4	Risques pour un mélange d'inverse-gamma et de gamma de paramètre d'échelle $\theta \geq 3$	64
2.5	Rapport des risques $\frac{R(\delta_{\psi^*}, \theta)}{R(\delta_0, \theta)}$ d'une Fisher décentrée de degrés de liberté 7 et 5 et de paramètre de décentralité $\theta \geq 5$	65
3.1	Bornes de l'EVM pour une chi-deux décentrée de $p = 6$ degrés de liberté et de paramètre de décentralité $\lambda \geq 0$	71
3.2	La différence $\hat{\lambda}_{\text{EVM}}(W) - (W - p - \lambda_0)_+$ pour une $\chi_p^2(\lambda)$ avec $\lambda \geq \lambda_0$ et $p = 2, 4, 6$	72

LISTE DES FIGURES

3.3	Rapport des risques $\frac{R(\hat{\lambda}_{\text{EVM},\lambda})}{R(\hat{\lambda}_+,\lambda)}$ d'une $\chi_p^2(\lambda)$ avec $\lambda \geq \lambda_0$ pour $p = 2, 4, 6$ et $\lambda_0 \in \{0, 2, 4, 6\}$	75
3.4	Risques quadratiques de $W - p$, $\hat{\lambda}_+$ et $\hat{\lambda}_\psi$ pour $W \sim \chi_p^2(\lambda)$ avec $\lambda \geq \lambda_0$ et $p = 4, 6$	77
3.5	Densités <i>a posteriori</i> pour une $\chi_p^2(\lambda)$ avec les observations $x \in \{0, 1; 1; 3; 5\}$, $p \in \{2, 4\}$ et $b \in \{0, 01; 0, 5; 1\}$	81
3.6	Risques quadratiques pondérés de $W - p$ et $(W - p)_+$ pour $W \sim \chi_p^2(\lambda)$ avec $\lambda \geq 0$ et $p \in \{2, 4, 6\}$	83
3.7	Densités prédictives d'une $\chi_4^2(\lambda)$ pour les observations $x \in \{0, 1; 1; 3; 5\}$ et $b \in \{0, 01; 0, 5; 1\}$	85

Abréviations et notations

$B(\alpha, \beta)$: fonction bêta pour les paramètres α, β .

$B_x(\alpha, \beta)$: fonction bêta incomplète de paramètres α et β où on intègre de 0 à x .

$\mathcal{B}(\alpha, \beta)$: loi bêta de paramètres de forme α, β .

$\mathcal{B}_{\text{II}}(\alpha, \beta)$: loi bêta de type II de paramètres de forme α, β .

$\mathcal{B}(\alpha, \beta; \theta)$: loi bêta décentrée de paramètres de forme α, β et de décentralité θ .

$b(\theta)$: biais.

$\mathcal{BN}(a, p)$: loi binomiale négative de paramètres a et p .

$\chi_p^2(\theta)$: loi du chi-deux à p degrés de liberté et de paramètre de décentralité θ .

$\text{Exp}(\beta, \theta)$: loi exponentielle de paramètres d'échelle β et de position θ .

$f(\cdot)$: densité marginale du modèle de mélange.

$F(\cdot)$: fonction de répartition du modèle de mélange.

$\bar{F}(\cdot)$: fonction de survie du modèle de mélange.

$f_k(\cdot)$: densité conditionnelle au paramètre de mélange K .

$F_k(\cdot)$: fonction de répartition conditionnelle au paramètre de mélange K .

$p_\theta(\cdot)$: densité du paramètre de mélange K .

${}_pF_q(\boldsymbol{\alpha}; \boldsymbol{\beta}; \cdot)$: fonction hypergéométrique généralisée de paramètres $\boldsymbol{\alpha} \in \mathbb{R}^p, \boldsymbol{\beta} \in \mathbb{R}^q$.

$(\cdot)_k$: symbole de Pochhammer, factorielle ascendante de k termes.

$I_\nu(\cdot)$: fonction de Bessel modifiée de première espèce d'ordre ν .

$Q_p(\cdot)$: rapport des fonctions de Bessel modifiées I_p et I_{p-1} .

$\mathbb{E}_\theta(\cdot)$: espérance du modèle de mélange sachant le paramètre θ .

ABRÉVIATIONS ET NOTATIONS

$\mathbb{E}_\theta^K(\cdot)$: espérance du paramètre de mélange K sachant le paramètre θ .

EVM : estimateur de vraisemblance maximale.

ESB : estimateur sans biais.

$F_{m,n}$: loi de Fisher à m et n degrés de liberté.

$\Gamma(\alpha)$: fonction gamma de paramètre α .

$\gamma_x(\alpha)$: gamma incomplète de paramètre α où on intègre de 0 à x .

$\mathcal{G}(\alpha, \beta)$: loi gamma de paramètres de forme α et d'échelle β .

$\mathcal{IG}(\alpha, \beta)$: loi inverse-gamma de paramètres de forme α et d'échelle β .

$\mathbb{1}_{[a,b]}(\cdot)$: fonction indicatrice pour l'intervalle $[a, b]$.

$\mathcal{I}(\cdot)$: information de Fisher.

$\text{Méd}_\theta(\cdot)$: médiane sachant le paramètre θ .

MEE : meilleur estimateur équivariant.

$\mathcal{N}(\mu, \sigma^2)$: loi normale de moyenne μ et de variance σ^2 .

$\phi(\cdot)$: densité d'une $\mathcal{N}(0, 1)$.

$\Phi(\cdot)$: fonction de répartition d'une $\mathcal{N}(0, 1)$.

$\mathcal{N}_m(\mu, \Sigma)$: loi normale multivariée de moyenne $\mu \in \mathbb{R}^m$ et de matrice de variance-covariance $\Sigma \in \mathbb{R}^m \times \mathbb{R}^m$ définie positive.

$\mathcal{LN}(\mu, \sigma^2)$: loi log-normale de paramètres μ et σ^2 .

$\mathcal{P}(\lambda)$: loi de Poisson de moyenne λ .

\propto : proportionnel à.

RVM : rapport de vraisemblance monotone.

$R(\delta, \theta)$: fonction de risque de l'estimateur δ par rapport au paramètre θ .

resp. : respectivement.

v.a. : variable aléatoire.

$\text{Var}_\theta(\cdot)$: variance par rapport au paramètre θ .

$\pi(\cdot)$: loi *a priori*.

$\pi(\cdot \mid X = x)$: loi *a posteriori* sachant l'observation x .

$f_{Y|X=x}(\cdot)$: densité prédictive de Y sachant les données observées x .

Introduction

Depuis la fin du 19^e siècle, la statistique a pu être appliquée dans une panoplie de domaines requérant de l'inférence ou une prise de décision. On peut notamment penser à la démographie, la biologie et la médecine, pour ne nommer que ceux-ci. De plus, grâce à l'avènement de l'ordinateur moderne et de l'informatique, initiés par Turing à partir 1936 [50], le traitement d'un grand nombre de données ainsi que des analyses complexes nécessitant de la puissance de calcul ont pu être effectués. Ainsi, en ce début du 21^e siècle, la statistique occupe une place importante dans la littérature scientifique.

Cette hausse d'intérêt pour la statistique et la modélisation a poussé la théorie plus loin afin de répondre aux problématiques complexes qu'on retrouve en pratique. Par exemple, beaucoup de jeux de données ayant des structures compliquées sont mal modélisés dans un cadre paramétrique simple : une solution prometteuse est l'étude des modèles de mélange comme ceux-ci offrent beaucoup de flexibilité. Leur étude a été introduite à la fin du 19^e siècle avec Pearson [39] qui modélise des données dites anormales à l'aide d'un mélange de deux lois normales. Toutefois, l'inférence pour une distribution de mélanges à l'aide de la méthode des moments entraînait un nombre de calculs important. Avec l'arrivée des ordinateurs de plus en plus puissants, l'estimation à l'aide de la vraisemblance avec l'algorithme itératif espérance-maximisation (EM) [6] a supplanté les techniques précédemment utilisées.

Comme les mélanges offrent beaucoup de flexibilité pour la modélisation de données, ceux-ci ont trouvé plusieurs domaines d'application dont la médecine [47] et le partitionnement de données [35], [36]. Cette solution a notamment été utilisée pour représenter des distributions bi-modales [16] et [33] ainsi que des distributions à queue longue [40] qui sont difficiles à approximer à l'aide d'un modèle paramétrique simple.

INTRODUCTION

Toutes ces applications ont mené à une étude approfondie des modèles de mélanges finis, particulièrement avec des lois normales et d'autres distributions de la famille exponentielle (voir par exemple [7], [34]). Le modèle présenté par Pigeon et Denuit [40] est particulièrement intéressant comme il combine un modèle de mélange fini avec un seuil aléatoire, c'est-à-dire un paramètre de mélange continu, ce qu'on retrouve peu en pratique [7].

Les situations où la variable d'intérêt n'est pas observable représente une autre application importante des distributions de mélange. Mandelkern [29] présente une situation commune en physique : les instruments mesurent un mélange de signal et de bruit de fond (*background*). Toutefois, certaines données sur le bruit de fond sont généralement disponibles comme il est possible d'effectuer des mesures sans signal. Cette information additionnelle peut ainsi être traduite comme une contrainte paramétrique afin d'être incorporée au modèle étudié.

L'étude de problèmes d'estimation en présence d'une restriction paramétrique a été introduite dans les années 50, s'intéressant particulièrement aux contraintes d'ordre (voir Robertson et coll. [44]). La statisticienne van Eeden, une précurseure du domaine, identifie dans son livre [51] plusieurs axes de recherche qu'on retrouve actuellement dans la littérature dont :

- la recherche d'estimateurs, préférablement admissibles, dominant d'autres inadmissibles ;
- la recherche d'estimateurs validant certains critères souhaitables tels que la minimaxité et l'admissibilité.

Ces thèmes seront aussi touchés dans les chapitres 2 et 3 de ce mémoire. Elle note aussi que le gain au niveau du risque d'un problème d'estimation sous contrainte n'a pas été un thème beaucoup abordé d'un point de vue théorique, bien que plusieurs résultats numériques se retrouvent dans la littérature. On obtient généralement une borne inférieure pour l'amélioration au niveau du risque, ce qui est la norme pour les problèmes d'estimation sous contrainte semi-bornée [51]. Dans ce mémoire, plusieurs graphiques de risques seront associés aux différents exemples afin d'illustrer les gains potentiels. Les articles de revue de Marchand et Strawderman [56], qui présente plusieurs techniques utiles pour l'estimation ponctuelle, et de Mandlkern [29], qui touche des intervalles de confiance, sont d'autres bonnes références sur le sujet.

INTRODUCTION

Ce mémoire se sépare en trois chapitres. Dans un premier temps, plusieurs notions préliminaires utiles pour l’obtention des principaux résultats sont présentées. On débute avec une brève présentation des distributions de mélanges telles que les chi-deux et Fisher décentrées. Ensuite, on discute de quelques ordres stochastiques dont le rapport de vraisemblance monotone (RVM), un élément clé des démonstrations du Chapitre 2. Quelques résultats originaux sont aussi obtenus pour des modèles de mélange. La section suivante introduit la théorie de la décision statistique qui permet notamment de comparer différents estimateurs entre eux. On conclut le Chapitre 1 avec une présentation de la méthode de Kubokawa. Celle-ci repose sur une astuce permettant de transformer la différence des fonctions de perte en intégrale et ainsi de déduire des conditions de dominance. Un pendant discret est aussi développé.

Dans le Chapitre 2, on présente des résultats originaux de dominance sous la perte quadratique pour des modèles de mélange continu. Sous certaines hypothèses assez générales telle que le RVM, on obtient une classe d’estimateurs lisses appartenant à l’espace paramétrique, et ce, même lorsque ce dernier est borné inférieurement. Les développements présentés sont une généralisation des résultats obtenus par Kubokawa, Marchand et Strawderman [23]. Parmi les applications discutées dans ce chapitre, on retrouve notamment les chi-deux et Fisher décentrées ainsi que la loi du coefficient de détermination expérimental R^2 en présence d’un modèle normal, un mélange de lois gamma ainsi qu’un modèle similaire à celui décrit par Pigeon et Denuit [40] avec un seuil aléatoire. De plus, on compte plusieurs graphiques des fonctions de risques tout au long du chapitre.

Finalement, le Chapitre 3 touche plus en profondeur à l’estimation du paramètre de décentralité d’une loi chi-deux décentrée. On présente premièrement plusieurs propriétés de la fonction de Bessel modifiée, avec laquelle on exprime la densité. Ensuite, on reprend quelques résultats de Saxena et Alam [46] sur l’estimateur de vraisemblance maximale (EVM) pour le problème d’estimation sans contrainte en plus de les généraliser pour un espace paramétrique borné inférieurement. De plus, quelques simplifications sont apportées à leur démonstration. Ainsi, on est en mesure de comparer les fonctions de risque de divers estimateurs et déduire des résultats de dominance sous la perte quadratique. On conclue ce chapitre avec les inférences bayésienne et prédictive sous un *a priori* gamma ainsi que l’identification d’un estimateur minimax.

Chapitre 1

Notions préliminaires et premiers résultats

On présente dans ce chapitre les notions sur lesquelles se basent les résultats et développements de ce mémoire. On discute premièrement des mélanges de lois de probabilité, le type de distributions traitées dans ces travaux. On y retrouve notamment les lois chi-deux et Fisher décentrées parmi d'autres. Dans la section suivante, on discute d'ordonnancements stochastiques, des critères qui assurent certaines propriétés mathématiques intéressantes. Le rapport de vraisemblance monotone (RVM), qui a de nombreuses applications dont les tests d'hypothèses (voir Chapitre 3 de [26]), sera un élément clé des démonstrations du Chapitre 2. À la section 1.3, on introduit les bases de la théorie de la décision statistique en plus d'explorer deux critères d'optimalité : l'équivariance et le critère minimax. On conclut ce chapitre avec une présentation de la méthode de Kubokawa qui permet d'exprimer la différence de fonctions de pertes comme une intégrale. Cette technique est particulièrement utile lorsqu'on souhaite améliorer un estimateur cible telle qu'une règle minimax.

1.1 Mélanges de lois de probabilité

La notion de mélange est généralement utilisée lorsque les paramètres suivent eux aussi une loi de probabilité, ce qu'on retrouve notamment sous le paradigme bayésien.

1.1. MÉLANGES DE LOIS DE PROBABILITÉ

Dans ce cas, il est possible d'exprimer le mélange comme un modèle hiérarchique. On a donc, par exemple,

$$\begin{aligned} X \mid K = k &\sim f_k \\ K &\sim p. \end{aligned} \tag{1.1}$$

On considère ici que K , le paramètre de mélange, est une variable aléatoire (v.a.) univariée. L'exemple le plus simple de mélange est lorsque deux populations distinctes se fusionnent. Les distributions tronquées où le seuil est aléatoire est l'un des modèles les plus simples où le paramètre de mélange est continu. Un autre exemple simple de modèle de mélange est le cas dégénéré : exactement un poids de la combinaison convexe est 1 et tous les autres sont nuls.

Dans le cas univarié, les mélanges se séparent en deux grandes catégories : ceux avec un paramètre de mélange continu qu'on appelle indénombrables et les autres où K est une v.a. discrète qui sont dits dénombrables.

1.1.1 Mélanges indénombrables

On dit d'une loi de probabilité qu'elle est un mélange indénombrable de lois lorsque sa fonction de densité peut s'écrire

$$f(x) = \int p(k) f_k(x) dk. \tag{1.2}$$

On reconnaît la représentation (1.1) où K est une v.a. continue. Ce type de mélange comprend notamment la binomiale négative et la bêta de type II [14].

Binomiale négative

La loi binomiale négative surgit lors d'un processus de comptage. Cette distribution, aussi appelée loi de Pòlya dans la littérature, est souvent utilisée pour modéliser des données dites surdispersées (« *overdispersed* »), où la loi de Poisson n'est pas appropriée puisque la variance échantillonnale est supérieure à sa moyenne. Sa représentation de mélange apparaît notamment dans le modèle Poisson-gamma en statistique bayésienne. Le livre de Johnson et coll. [13] offre une présentation plus

1.1. MÉLANGES DE LOIS DE PROBABILITÉ

complète de cette distribution.

Définition 1.1.1. *On dit que X suit une loi binomiale négative de paramètres $p \in (0, 1)$ et $a > 0$ si sa fonction de masse s'écrit*

$$f(x) = \frac{\Gamma(a+x)}{\Gamma(a)x!} (1-p)^x p^a \mathbb{1}_{\mathbb{N}}(x). \quad (1.3)$$

On note $X \sim \mathcal{BN}(a, p)$.

Ainsi, on peut facilement démontrer les lemmes suivants.

Lemme 1.1.1. *Soit $X \sim \mathcal{BN}(a, p)$. Alors, $\mathbb{E}_p(X) = a \frac{1-p}{p}$ et $\text{Var}_p(X) = a \frac{1-p}{p^2}$.*

Lemme 1.1.2. *Soit $X \sim \mathcal{BN}\left(a, \frac{1}{\theta+1}\right)$ avec $a, \theta > 0$. Une représentation équivalente à (1.3) est*

$$\begin{aligned} X \mid K = k &\sim \mathcal{P}(k) \\ K &\sim \mathcal{G}(a, \theta) \end{aligned} \quad (1.4)$$

avec $f_k(x) = \frac{k^x e^{-k}}{x!} \mathbb{1}_{\mathbb{N}}(x)$ et $p_\theta(k) = \frac{k^a e^{-\frac{k}{\theta}}}{\Gamma(a)\theta^a} \mathbb{1}_{(0, \infty)}(k)$.

Démonstration. Lorsqu'on reprend le mélange (1.4), on obtient

$$\begin{aligned} f(x; \theta) &= \int_0^\infty f_k(x) p_\theta(k) dk = \frac{1}{x! \Gamma(a) \theta^a} \mathbb{1}_{\mathbb{N}}(x) \int_0^\infty k^{a+x-1} e^{-k(1+1/\theta)} dk \\ &= \frac{\Gamma(a+x) \left(\frac{\theta}{1+\theta}\right)^{a+x}}{x! \Gamma(a) \theta^a} \mathbb{1}_{\mathbb{N}}(x) = \frac{\Gamma(a+x)}{x! \Gamma(a)} \left(\frac{1}{1+\theta}\right)^a \left(\frac{\theta}{1+\theta}\right)^x \mathbb{1}_{\mathbb{N}}(x) \end{aligned}$$

et on retrouve l'équation (1.3). □

Bêta de type II

La loi bêta de type II, à valeurs dans \mathbb{R}_+ , est obtenue à l'aide de la transformation $T : u \mapsto u/(1-u)$ sur la loi bêta standard. Cette distribution, aussi appelée bêta prime dans la littérature, permet de modéliser une situation en terme de chances (*odds*). Le livre de Johnson et coll. [15] offre une présentation plus complète de cette distribution.

1.1. MÉLANGES DE LOIS DE PROBABILITÉ

Définition 1.1.2. On dit que X suit une loi bêta de type II de paramètres $\alpha, \beta > 0$ si

$$X \stackrel{d}{=} \frac{Y}{Z} \quad (1.5)$$

avec $Y \sim \mathcal{G}(\alpha, 1)$ et $Z \sim \mathcal{G}(\beta, 1)$ indépendants. On note $X \sim \mathcal{B}_{\text{II}}(\alpha, \beta)$.

Lemme 1.1.3. Soit $X \sim \mathcal{B}_{\text{II}}(\alpha, \beta)$ pour $\alpha, \beta > 0$. Alors sa fonction de densité s'écrit

$$f(x) = \frac{x^{\alpha-1}(1+x)^{-\alpha-\beta}}{\text{B}(\alpha, \beta)} \mathbb{1}_{(0, \infty)}(x) \quad (1.6)$$

avec $\text{B}(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$. De plus, on a que $\mathbb{E}(X) = \frac{\alpha}{\beta-1}$ pour $\beta > 1$ et que $\text{Var}(X) = \frac{\alpha+\beta-1}{(\beta-2)(\beta-1)^2}$ pour $\beta > 2$.

De l'équation (1.5), on trouve la représentation de mélange suivante.

Lemme 1.1.4. Soit $X \sim \mathcal{B}_{\text{II}}(\alpha, \beta)$ avec $\alpha, \beta > 0$. Une représentation équivalente à (1.6) est

$$\begin{aligned} X \mid K = k &\sim \mathcal{G}(\alpha, 1/k) \\ K &\sim \mathcal{G}(\beta, 1). \end{aligned}$$

1.1.2 Mélanges dénombrables

On dit d'une loi de probabilité qu'elle est un mélange dénombrable lorsque que sa distribution est une combinaison convexe de fonctions de densité. Supposons que la densité de X peut s'écrire

$$f(x) = \sum_k p(k) f_k(x). \quad (1.7)$$

On reconnaît la représentation (1.1) où K est une v.a. discrète de fonction de masse p . Ce type de mélanges regroupe plusieurs lois de probabilité dont les chi-deux, bêta et Fisher décentrées [15]. De plus, on trouve que le coefficient de détermination R^2 sous un modèle normal ainsi que la transformation $\frac{R^2}{1-R^2}$ se décomposent aussi comme des mélanges dénombrables.

1.1. MÉLANGES DE LOIS DE PROBABILITÉ

Loi chi-deux décentrée

La chi-deux décentrée est une loi fondamentale pour les formes quadratiques apparaissant notamment dans la théorie des modèles linéaires [37]. Considérons un échantillon X_1, \dots, X_p indépendantes telles que $X_i \sim \mathcal{N}(\mu_i, 1)$ pour $i \in \{1, \dots, p\}$. On obtient que

$$Y \stackrel{d}{=} \sum_{i=1}^p X_i^2$$

est distribué selon une chi-deux décentrée à n degrés de liberté et de paramètre de décentralité $\theta = \sum_{i=1}^n \mu_i^2$. Pour p , un réel positif, on obtient l'extension suivante.

Définition 1.1.3. *On dit que X suit une loi chi-deux à $p > 0$ degrés de liberté avec paramètre de décentralité $\theta > 0$ si sa fonction de densité s'écrit*

$$f(x) = \frac{1}{2} e^{-\frac{x+\theta}{2}} \left(\frac{x}{\theta}\right)^{\frac{p-2}{4}} I_{\frac{p}{2}-1}(\sqrt{\theta x}) \mathbb{1}_{(0,\infty)}(x) \quad (1.8)$$

où $I_\nu(y) = \left(\frac{y}{2}\right)^\nu \sum_{k \geq 0} \frac{1}{\Gamma(\nu+k+1)} \frac{\left(\frac{y^2}{4}\right)^k}{k!}$ est la fonction de Bessel modifiée d'ordre ν . On note $X \sim \chi_p^2(\theta)$. Elle est dite centrée lorsque sa fonction de densité s'écrit

$$f(x) = \frac{x^{\frac{p}{2}-1} e^{-\frac{x}{2}}}{\Gamma(p/2) 2^{\frac{p}{2}}}.$$

Lemme 1.1.5. *Soit $X \sim \chi_p^2(\theta)$. Alors $\mathbb{E}_\theta(X) = p + \theta$ et $\text{Var}_\theta(X) = 2(p + 2\theta)$.*

Pour une présentation plus complète de cette distribution, on peut se référer au livre de Johnson et coll. [13]. De plus, en réarrangeant les termes de l'équation (1.8), on remarque que celle-ci vérifie la représentation (1.7) et on obtient le lemme suivant.

Lemme 1.1.6. *On a qu'une représentation équivalente à (1.8) est*

$$\begin{aligned} X \mid K = k &\sim \chi_{p+2k}^2(0) \\ K &\sim \mathcal{P}(\theta/2). \end{aligned} \quad (1.9)$$

Démonstration. On remarque premièrement que la densité de la loi de $X \mid K = k$

1.1. MÉLANGES DE LOIS DE PROBABILITÉ

s'écrit

$$f_k(x) = \frac{e^{-\frac{x}{2}} x^{p/2+k-1}}{\Gamma(p/2) 2^{p/2}} \mathbb{1}_{(0,\infty)}(x).$$

On reprend (1.9) et on obtient que la densité marginale est

$$\begin{aligned} f(x) &= \sum_{k \geq 0} \frac{(\theta/2)^k e^{-\theta/2}}{k!} \frac{e^{-\frac{x}{2}} x^{p/2+k-1}}{\Gamma(p/2 + k) 2^{p/2+k}} \mathbb{1}_{(0,\infty)}(x) \\ &= \frac{1}{2} e^{-\frac{x+\theta}{2}} (x/2)^{p/2-1} \mathbb{1}_{(0,\infty)}(x) \sum_{k \geq 0} \frac{(x\theta/4)^k}{\Gamma(p/2 + k) k!} \\ &= \frac{1}{2} e^{-\frac{x+\theta}{2}} \frac{(x/2)^{p/2-1}}{\left(\frac{\sqrt{x\theta}}{2}\right)^{p/2-1}} I_{p/2-1}(\sqrt{x\theta}) \mathbb{1}_{(0,\infty)}(x) \end{aligned}$$

et on reconnaît la densité d'une chi-deux décentrée donnée à l'équation (1.8). □

Mélanges de bêta de type II

On s'intéresse ici aux lois de probabilité dont la densité est de la forme

$$f(x; \theta) = \sum_{k \geq 0} \frac{x^{\alpha+k-1} (1+x)^{-\alpha-\beta}}{B(\alpha+k, \beta)} \mathbb{1}_{(0,\infty)}(x) p_\theta(k)$$

pour des paramètres $\alpha, \beta > 0$. En raison de la construction de la loi bêta de type II, la loi Fisher décentrée et la loi de la transformation du coefficient de détermination expérimental $R^2/(1-R^2)$, où les observations sont issues d'une loi normale multivariée, comptent parmi les mélanges admettant cette représentation.

Loi de Fisher décentrée

La loi de Fisher décentrée surgit lorsqu'on considère le quotient de deux lois chi-deux indépendantes dont une est centrée, une caractérisation très similaire à celle de la loi de Fisher standard. Le livre de Jonhson et coll. [13] offre une présentation plus complète de cette distribution.

Définition 1.1.4. *On dit que X suit une loi de Fisher décentrée de degrés de liberté*

1.1. MÉLANGES DE LOIS DE PROBABILITÉ

$\nu_1, \nu_2 > 0$ avec paramètre de décentralité $\theta \geq 0$ si

$$X \stackrel{d}{=} \frac{Y/\nu_1}{Z/\nu_2} \quad (1.10)$$

avec $Y \sim \chi_{\nu_1}^2(\theta)$ et $Z \sim \chi_{\nu_2}^2(0)$ indépendants. On note $X \sim F_{\nu_1, \nu_2}(\theta)$ et elle est dite centrée lorsque $\theta = 0$.

En raison de l'équation (1.10), on note plusieurs similitudes avec la chi-deux décentrée, notamment en ce qui a trait à la représentation de mélange.

Lemme 1.1.7. *Soit $X \sim F_{\nu_1, \nu_2}(\theta)$. Alors, sa fonction de densité s'écrit*

$$f(x) = \sum_{k \geq 0} \frac{e^{-\frac{\theta}{2}} (\theta/2)^k}{k! B(k + \nu_1/2, \nu_2/2)} \left(\frac{\nu_1}{\nu_2} \right)^{\frac{\nu_1}{2} + k} \left(\frac{\nu_2}{\nu_2 + \nu_1 x} \right)^{\frac{\nu_1 + \nu_2}{2} + k} x^{\frac{\nu_1}{2} + k - 1} \mathbb{1}_{(0, \infty)}(x). \quad (1.11)$$

De plus, on a que,

$$\mathbb{E}_\theta(X) = \begin{cases} \frac{\nu_2(\nu_1 + \theta)}{\nu_1(\nu_2 - 2)} & \text{si } \nu_2 > 2 \\ \text{pas définie} & \text{sinon,} \end{cases}$$

et que

$$\text{Var}_\theta(X) = \begin{cases} 2 \left(\frac{\nu_2}{\nu_1} \right)^2 \frac{(\nu_1 + \theta)^2 + (\nu_2 - 2)(\nu_1 + 2\theta)}{(\nu_2 - 2)^2(\nu_2 - 4)} & \text{si } \nu_2 > 4 \\ \text{pas définie} & \text{sinon.} \end{cases}$$

On remarque aussi que l'équation (1.11) vérifie la représentation (1.7) et on obtient le lemme suivant. En raison du lien entre les lois de Fisher centrée et bêta de type II, on trouve deux représentations distinctes.

Lemme 1.1.8. *Les représentations suivantes sont équivalentes à (1.11)*

$$\begin{aligned} \frac{\nu_1}{\nu_2} X \mid K = k &\sim \mathcal{B}_{\text{II}} \left(\frac{\nu_1}{2} + k, \frac{\nu_2}{2} \right) & X \mid K = k &\sim F_{\nu_1 + 2k, \nu_2}(0) \\ K &\sim \mathcal{P}(\theta/2) & K &\sim \mathcal{P}(\theta/2). \end{aligned}$$

Transformation du coefficient de détermination expérimental $R^2/(1-R^2)$

Soit $\mathbf{V} = (V_1, \mathbf{V}_2)^T$, avec $\mathbf{V}_2 = (V_2, \dots, V_m)^T$, issu d'une distribution normale multivariée de moyenne $\mu \in \mathbb{R}^m$ et matrice de variances-covariances Σ définie positive

1.1. MÉLANGES DE LOIS DE PROBABILITÉ

connues. Posons ρ^2 le coefficient de détermination entre les variables V_1 et \mathbf{V}_2 prenant des valeurs entre 0 et 1. Ce dernier est le carré du coefficient de corrélation ρ . On retrouve une présentation plus détaillée dans le livre de Muirhead [37].

Pour une taille d'échantillon $n + 1$, le coefficient de détermination expérimental R^2 , une quantité utilisée dans de nombreux contextes dont les modèles linéaires et l'analyse multivariée [37], est l'estimateur de vraisemblance maximale (EVM) de ρ^2 . Muirhead [38] présente plusieurs arguments, dont la stabilité numérique, pour privilégier l'estimation de la transformation $\theta = \frac{\rho^2}{1-\rho^2}$ basée sur $X \stackrel{d}{=} \frac{R^2}{1-R^2}$ plutôt que celle de ρ^2 basée directement sur $Y \stackrel{d}{=} R^2$.

Lemme 1.1.9 (Théorème 5.2.5 de [37]). *Pour m variables et un échantillon de taille $n + 1$, on a que $X \stackrel{d}{=} \frac{R^2}{1-R^2}$ admet la représentation suivante :*

$$\begin{aligned} X \mid K = k &\sim \mathcal{B}_{\text{II}}\left(\frac{m-1}{2} + k, \frac{n-m+1}{2}\right) \\ K &\sim \mathcal{BN}(n/2, (1+\theta)^{-1}). \end{aligned} \tag{1.12}$$

Un autre argument mis de l'avant par Muirhead est que les moments de X sont plus faciles à représenter que ceux de Y .

Lemme 1.1.10. *Soit X issu du modèle (1.12). Alors, on a que*

$$\mathbb{E}_{\theta}(X) = \begin{cases} \frac{m-1+n\theta}{n-m-1} & \text{si } n-m > 1 \\ \text{pas définie} & \text{sinon,} \end{cases}$$

et que

$$\text{Var}_{\theta}(X) = \begin{cases} \frac{2n\theta(2+\theta)}{(n-m-1)^2} + \frac{4n\theta(1+\theta)+2n^2\theta^2+2(m-1)(n-2)}{(n-m-3)(n-m-1)^2} & \text{si } n-m > 3 \\ \text{pas définie} & \text{sinon.} \end{cases}$$

Démonstration. À l'aide du Lemme 1.1.3 et de la représentation de mélange, on trouve, pour $n-m > 1$, que

$$\mathbb{E}_{\theta}(X) = \mathbb{E}_{\theta}^K \mathbb{E}(X \mid K = k) = \mathbb{E}_{\theta}^K \left[\frac{\frac{m-1}{2} + K}{\frac{n-m+1}{2} - 1} \right] = \frac{m-1+n\theta}{n-m-1}$$

1.1. MÉLANGES DE LOIS DE PROBABILITÉ

et, pour $n - m > 3$, que

$$\begin{aligned}\text{Var}_\theta(X) &= \text{Var}_\theta^K \mathbb{E}(X \mid K = k) + \mathbb{E}_\theta^K \text{Var}(X \mid K = k) \\ &= \text{Var}_\theta^K \left[\frac{m-1+2K}{n-m-1} \right] + \mathbb{E}_\theta^K \left[8 \left(\frac{K^2 + K \frac{n-m-3}{2} + \frac{(m-1)(n-2)}{4}}{(n-m-3)(n-m-1)^2} \right) \right] \\ &= \frac{2n\theta(2+\theta)}{(n-m-1)^2} + \frac{4n\theta(1+\theta) + 2n^2\theta^2 + 2(m-1)(n-2)}{(n-m-3)(n-m-1)^2},\end{aligned}$$

ce qui démontre le résultat. \square

Mélanges de bêta de type I

Définition 1.1.5. *On dit que X suit une loi bêta de type I de paramètres $\alpha, \beta > 0$ lorsque sa densité s'écrit*

$$f(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{\text{B}(\alpha, \beta)} \mathbb{1}_{(0,1)}(x).$$

On note $X \sim \mathcal{B}(\alpha, \beta)$.

On s'intéresse ici aux lois de probabilité dont la densité est de la forme

$$f(x; \theta) = \sum_{k \geq 0} \frac{x^{\alpha+k-1}(1-x)^{\beta-1}}{\text{B}(\alpha+k, \beta)} \mathbb{1}_{(0,1)}(x) p_\theta(k)$$

pour des paramètres $\alpha, \beta > 0$. Marchand [53] fait une présentation plus détaillée de ce type de mélanges. En raison des moments de la loi bêta, ceux des mélanges considérés dans cette sous-section n'ont pas de forme close, mais peuvent être représentés à l'aide des fonctions hypergéométriques généralisées (présentées à la Définition 3.3.1).

Bêta décentrée

La loi bêta décentrée survient lorsqu'on considère un quotient de deux chi-deux, une décentré et l'autre centrée, de sorte que le rapport varie entre 0 et 1. Le livre de Johnson et coll. [15] offre une présentation plus complète de cette distribution.

Définition 1.1.6. *On dit que X suit une loi bêta décentrée de paramètres $\alpha, \beta > 0$*

1.2. ORDONNANCEMENT STOCHASTIQUE

avec paramètre de décentralité $\theta \geq 0$ lorsque

$$X \stackrel{d}{=} \frac{Y}{Y + Z} \quad (1.13)$$

où $Y \sim \chi_\alpha^2(\theta)$ et $Z \sim \chi_\beta^2(0)$. On note $X \sim \mathcal{B}(\alpha/2, \beta/2; \theta)$.

En raison de sa construction, la représentation de mélange de la bêta décentrée est très similaire à celle de la Fisher décentrée.

Lemme 1.1.11. *On a qu'une représentation équivalente à (1.13) est*

$$\begin{aligned} X \mid K = k &\sim \mathcal{B}(k + \alpha/2, \beta/2) \\ K &\sim \mathcal{P}(\theta/2). \end{aligned}$$

Coefficient de détermination expérimental

Tel que mentionné précédemment, l'estimation des moments du coefficient de détermination expérimental est complexifiée en raison des moments de la loi bêta. Marchand [54] traite de l'estimation du coefficient de corrélation à l'aide d'une fonction affine de R^2 ainsi qu'avec une approche bayésienne.

Lemme 1.1.12. *Pour un échantillon de taille $n + 1$ et m variables, on a que $Y \stackrel{d}{=} R^2$ admet la représentation suivante :*

$$\begin{aligned} Y \mid K = k &\sim \mathcal{B}\left(\frac{m-1}{2} + k, \frac{n-m+1}{2}\right) \\ K &\sim \mathcal{BN}(n/2, (1 + \theta)^{-1}). \end{aligned}$$

1.2 Ordonnancement stochastique

Les ordres stochastiques représentent divers critères afin de qualifier qu'une variable aléatoire est « plus grande » qu'une autre, « plus variable » qu'une autre, etc. Il existe plusieurs relations d'ordre pour les variables aléatoires univariées dont la croissance stochastique, les ordres du rapport de vraisemblance et du taux de défaillance. Dans cette section, seuls les deux premiers ordonnancements mentionnés sont abordés pour des familles de distributions non négatives. Marshall et coll. [32] offrent une présentation plus complète des divers ordres stochastiques.

1.2. ORDONNANCEMENT STOCHASTIQUE

On s'intéresse premièrement à l'ordonnancement stochastique comme celui-ci assure certaines propriétés, particulièrement en ce qui concerne la caractérisation d'espérances. On discute ensuite du rapport de vraisemblance monotone (RVM), un critère plus utilisé en pratique, notamment en raison de son lien avec les tests d'hypothèses tel que présenté par Lehmann et Romano [26].

1.2.1 Croissance stochastique

Définition 1.2.1. Soit $X \sim f_\theta$, une v.a. sur $\mathbb{R}_+ \cup \{0\}$. On dit que la famille de densités $\{f_\theta : \theta \in \Theta\}$ a un ordre stochastiquement croissant (resp. décroissant) en X avec paramètre θ si la fonction de survie $\bar{F}_\theta(x) = \mathbb{P}_\theta(X > x)$ est croissante (resp. décroissante) en θ pour tout $x \geq 0$.

Autrement dit, pour une variable stochastiquement croissante, moins θ est grand, plus la masse est concentrée pour des petites valeurs de son domaine.

Exemple 1.2.1. Voici quelques familles de distributions f_θ avec un ordre stochastiquement croissant. Posons f_θ , la densité de X .

1. Soit Y une v.a. continue de densité $f(t) > 0$ lorsque $t \in (a, b)$ et 0 sinon. Soient $\theta \in (a, b)$ et $X \stackrel{d}{=} Y \mid Y \leq \theta$. Alors f_θ est stochastiquement croissante en X avec paramètre θ , car pour $x < \theta$,

$$\bar{F}_\theta(x) = \mathbb{P}_\theta(X > x) = \frac{\mathbb{P}(x < Y \leq \theta)}{\mathbb{P}(Y \leq \theta)} = 1 - \frac{\mathbb{P}(Y \leq x)}{\mathbb{P}(Y \leq \theta)},$$

ce qui est bien croissant en θ .

2. Considérons $X \sim \mathcal{P}(\theta)$. Alors f_θ est stochastiquement croissante en X avec paramètre θ , car pour tout $x \in \mathbb{N}$,

$$\frac{\partial}{\partial \theta} \bar{F}_\theta(x) = e^{-\theta} \left(\sum_{i>x} \frac{\theta^{i-1}}{(i-1)!} - \frac{\theta^i}{i!} \right) = \frac{e^{-\theta} \theta^x}{x!} \geq 0,$$

ce qui est bien croissant en θ .

Lemme 1.2.2 (Proposition 17.A.2 [32]). Soit $X \sim f_\theta$ avec $\theta \in \mathbb{R}_+$. Les conditions suivantes sont équivalentes :

1.2. ORDONNANCEMENT STOCHASTIQUE

- (i) la famille de densités f_θ a un ordre stochastiquement croissant en x avec paramètre θ ;
- (ii) $\mathbb{E}_\theta(\phi(X))$ est croissante en θ pour toute fonction croissante ϕ si l'espérance est finie ;
- (iii) $\bar{G}_\theta(x)$ est croissante en θ , où $\bar{G}_\theta(\cdot)$ est la fonction de survie de $\phi(X)$ pour toute fonction strictement croissante ϕ .

Démonstration. On se base sur la preuve de [32]. Soient $\theta < \theta'$ et $x \in \mathbb{R}_+$. Posons $\phi^{-1}(x) = \inf\{y \in \mathbb{R} \mid \phi(y) \geq x\}$, l'inverse généralisée.

Si la condition (i) est vérifiée, on a que

$$\bar{G}_\theta(x) = \mathbb{P}_\theta(X > \phi^{-1}(x)) \leq \mathbb{P}_{\theta'}(X > \phi^{-1}(x)) = \bar{G}_{\theta'}(x)$$

et (i) \Rightarrow (iii). Soit (iii). Alors $0 \leq \bar{G}_{\theta'}(x) - \bar{G}_\theta(x)$. On a donc, lorsque X est continu, que $\int \bar{G}_\theta(x)dx \leq \int \bar{G}_{\theta'}(x)dx$ et on trouve, en interchangeant les intégrales, que

$$\int_{\mathbb{R}_+} \bar{G}_\theta(x)dx = \int_0^\infty \int_x^\infty g_\theta(k)dkdx = \int_0^\infty g_\theta(k) \int_0^k dx dk = \mathbb{E}_\theta(\phi(X))$$

où g_θ est la fonction de densité de $\phi(X)$. Pour X discret, on obtient le même résultat en remplaçant les intégrales par des sommes et on déduit que (iii) \Rightarrow (ii). Si la condition (ii) est vérifiée, on pose $\phi(x) = \mathbb{1}_{(a, \infty)}(x)$ et, pour $a > 0$, on a que

$$\mathbb{E}_\theta(\mathbb{1}_{(a, \infty)}(X)) = \mathbb{P}_\theta(a < X) = \bar{F}_\theta(a)$$

est croissant en θ et on trouve la condition (i). □

Exemple 1.2.3 (Suite de l'Exemple 1.2.1.1). *Comme la densité de $X \stackrel{d}{=} Y \mid Y \leq \theta$ est stochastiquement croissante en X avec paramètre θ , son espérance est croissante en θ .*

Bien que cette restriction soit peu sévère, travailler avec la fonction de survie n'est pas toujours commode puisqu'une somme ou une intégrale semi bornée est généralement impliquée. Le prochain ordonnancement requiert un rapport de densités, des quantités souvent bien plus simples à manipuler.

1.2. ORDONNANCEMENT STOCHASTIQUE

1.2.2 Rapport de vraisemblance monotone

Définition 1.2.2. Soit $X \sim f_\theta$. On dit que la famille de densités f_θ a un rapport de vraisemblance monotone (RVM) croissant (resp. décroissant) en X avec paramètre θ si, pour $\theta_1 < \theta_2$, le rapport $\frac{f_{\theta_2}(x)}{f_{\theta_1}(x)}$ est croissant (resp. décroissant) en x .

Exemple 1.2.4. Posons f_θ , la densité de X . Voici quelques familles de distributions f_θ avec un RVM croissant en X avec paramètre θ . Soient $\theta_1 < \theta_2$.

1. Considérons $Y \sim \mathcal{B}_{\text{II}}(\theta, b)$ avec b fixe et $\theta > 0$. Alors, pour $x > 0$, le rapport

$$\frac{f_{\theta_2}(x)}{f_{\theta_1}(x)} = \frac{x^{\theta_2-1}(1+x)^{-\theta_2-b}/\text{B}(\theta_2, b)}{x^{\theta_1-1}(1+x)^{-\theta_1-b}/\text{B}(\theta_1, b)} = \frac{\text{B}(\theta_1, b)}{\text{B}(\theta_2, b)} \left(\frac{x}{1+x}\right)^{\theta_2-\theta_1} \propto \left(\frac{x}{1+x}\right)^{\theta_2-\theta_1}$$

est croissant en x .

2. Considérons $X \sim \mathcal{P}(\theta)$ avec $\theta > 0$. Alors, pour $x \in \mathbb{N}$, le rapport

$$\frac{f_{\theta_2}(x)}{f_{\theta_1}(x)} = \frac{\theta_2^x e^{-\theta_2}/x!}{\theta_1^x e^{-\theta_1}/x!} = (\theta_2/\theta_1)^x e^{-(\theta_2-\theta_1)} \propto (\theta_2/\theta_1)^x$$

est croissant en x .

On remarque que la loi de Poisson satisfait les deux ordonnancements présentés dans cette section (voir l'Exemple 1.2.1.2). C'est en réalité le cas de toutes les distributions satisfaisant l'ordre du rapport de vraisemblance.

Lemme 1.2.5. Soit $X \sim f_\theta$. Supposons que f_θ a un RVM croissant (resp. décroissant) en X avec paramètre θ . Alors la famille de densités f_θ est stochastiquement croissante (resp. décroissante) en X avec paramètre θ .

Démonstration. Soient $\theta_1 < \theta_2$ et $x_1 < x_2$. Alors, on a que

$$\int_x^\infty f_{\theta_2}(x_2)f_{\theta_1}(x_1)dx_2 \geq \int_x^\infty f_{\theta_1}(x_2)f_{\theta_2}(x_1)dx_2 \Rightarrow \frac{\bar{F}_{\theta_2}(x)}{\bar{F}_{\theta_1}(x)} \geq \frac{f_{\theta_2}(x)}{f_{\theta_1}(x)} \quad (1.14)$$

et que

$$\int_{-\infty}^x f_{\theta_2}(x_2)f_{\theta_1}(x_1)dx_2 \geq \int_{-\infty}^x f_{\theta_1}(x_2)f_{\theta_2}(x_1)dx_2 \Rightarrow \frac{f_{\theta_2}(x)}{f_{\theta_1}(x)} \geq \frac{1 - \bar{F}_{\theta_2}(x)}{1 - \bar{F}_{\theta_1}(x)} \quad (1.15)$$

1.2. ORDONNANCEMENT STOCHASTIQUE

où \bar{F}_θ est la fonction de survie de X . En combinant les inégalités (1.14) et (1.15), on trouve que

$$\frac{\bar{F}_{\theta_2}(x)}{\bar{F}_{\theta_1}(x)} \geq \frac{1 - \bar{F}_{\theta_2}(x)}{1 - \bar{F}_{\theta_1}(x)} \Leftrightarrow \bar{F}_{\theta_2}(x) \geq \bar{F}_{\theta_1}(x). \quad \square$$

Les lois tronquées, qui jouent un rôle important dans ces travaux, satisfont aussi à cet ordonnancement.

Lemme 1.2.6. *Soit Y une variable aléatoire quelconque. On a que f_t , la densité de $X \stackrel{d}{=} Y \mid Y \leq t$, a un RVM croissant en X avec paramètre t .*

Démonstration. Soit $t' < t$. Alors, f_t s'écrit

$$f_t(x) = \frac{f(x)\mathbb{1}_{(-\infty, t]}(x)}{\mathbb{P}(X \leq t)}.$$

Pour $x \in (-\infty, t)$, on trouve

$$\frac{f_t(x)}{f_{t'}(x)} = \frac{\frac{f(x)\mathbb{1}_{(-\infty, t]}(x)}{\mathbb{P}(X \leq t)}}{\frac{f(x)\mathbb{1}_{(-\infty, t']}(x)}{\mathbb{P}(X \leq t')}} \propto \frac{\mathbb{1}_{(-\infty, t]}(x)}{\mathbb{1}_{(-\infty, t']}(x)} = \begin{cases} 1 & \text{si } x \leq t' \\ \infty & \text{si } t' < x \leq t. \end{cases}$$

Pour $x > t$, le rapport est indéterminé. On a donc le résultat. \square

Remarque 1.2.7. *En vertu des lemmes 1.2.6 et 1.2.5, les lois tronquées satisfont à l'ordre de la croissance stochastique, ce qui avait été présenté à l'Exemple 1.2.1.1.*

Pour le mélange de deux lois univariées, on peut déduire plusieurs propriétés à l'aide des RVM. Supposons le modèle suivant :

$$\begin{aligned} X \mid K = k &\sim f_k \\ K &\sim p_\theta \end{aligned} \tag{1.16}$$

où K est de support $\mathbb{K} \subset \mathbb{R}_+ \cup \{0\}$ avec $\{p_\theta \mid \theta \in \Theta\}$ une famille de densités ayant un RVM croissant en K avec paramètre θ et $\{f_k \mid k \in \mathbb{K}\}$ est une famille de densités. À l'aide de cette représentation, on peut démontrer que la marginale de X ainsi que la densité de la loi conditionnelle de $K \mid X \leq t$ satisfont à l'ordonnancement stochastique du rapport de vraisemblances. Ces lemmes sont particulièrement utiles lorsqu'on travaille avec des espérances conditionnelles.

1.2. ORDONNANCEMENT STOCHASTIQUE

Lemme 1.2.8 (Lemme 5.1 de [23]). *Dans le cadre du modèle (1.16), on suppose en plus que f_k a un RVM croissant en X avec paramètre k . Alors, la famille de densités $\{f(x; \theta) = \mathbb{E}_\theta^K(f_K(x)) \mid \theta \in \Theta\}$ a un RVM croissant en X avec paramètre θ .*

Démonstration. On se base sur la démonstration de [23]. Soient $\theta < \theta'$. Pour un x fixe, on a que

$$\frac{f(x; \theta')}{f(x; \theta)} = \frac{\mathbb{E}_\theta^K \left(\frac{p_{\theta'}(K)}{p_\theta(K)} f_K(x) \right)}{f(x; \theta)} = \mathbb{E}^L \left(\frac{p_{\theta'}(L)}{p_\theta(L)} \right)$$

où $L \sim f_l(x)p_\theta(l)/f(x; \theta)$. Or, pour $x < x'$, on a que

$$\frac{f_l(x')p_\theta(l)/f(x'; \theta)}{f_l(x)p_\theta(l)/f(x; \theta)} = \frac{f(x; \theta)}{f(x'; \theta)} \frac{f_l(x')}{f_l(x)} \propto \frac{f_l(x')}{f_l(x)}$$

est croissant en l puisque f_k a un RVM croissant en X avec paramètre k . Ainsi, la densité de L a un RVM croissant en L avec paramètre x . Comme on suppose que p_θ a un RVM croissant en K avec paramètre θ , le rapport $\frac{p_{\theta'}(L)}{p_\theta(L)}$ est croissant en L , d'où $\frac{f(x; \theta')}{f(x; \theta)}$ est croissant en x par la représentation (1.16). \square

Le lemme suivant, particulièrement utile pour la manipulation d'espérances conditionnelles, est un résultat original.

Lemme 1.2.9. *Soit le modèle (1.16) et posons $g_\theta(\cdot; t)$, la densité de $L \stackrel{d}{=} K \mid X \leq t$. Alors, pour une constante t telle que $\mathbb{P}_\theta(X \leq t) \neq 0$ pour tout θ , on a que g_θ a aussi un RVM croissant en L avec paramètre θ .*

Démonstration. Pour K une v.a. continue, posons F_k , la répartition de $X \mid K = k$. On a que

$$\mathbb{P}_\theta(L \leq k) = \mathbb{P}_\theta(K \leq k \mid X \leq t) = \frac{\mathbb{P}(X \leq t, K \leq k)}{\mathbb{P}_\theta(X \leq t)} = \frac{\int_{-\infty}^k F_y(t) p_\theta(y) dy}{\mathbb{P}_\theta(X \leq t)}.$$

En vertu du théorème fondamental du calcul intégral, on déduit

$$g_\theta(k; t) = \frac{d}{dk} \mathbb{P}_\theta(L \leq k) = \frac{F_k(t) p_\theta(k)}{\mathbb{P}_\theta(X \leq t)}.$$

1.3. THÉORIE DE LA DÉCISION STATISTIQUE

Ainsi, pour $\theta > \theta'$, on trouve que le quotient de densités

$$\frac{g_\theta(k; t)}{g_{\theta'}(k; t)} = \frac{F_k(t)p_\theta(k)/\mathbb{P}_\theta(X \leq t)}{F_k(t)p_{\theta'}(k)/\mathbb{P}_{\theta'}(X \leq t)} \propto \frac{p_\theta(k)}{p_{\theta'}(k)} \quad (1.17)$$

est croissant en k par hypothèse. Pour K discret, on reprend l'équation (1.17) et on trouve le résultat. \square

On illustre les résultats plus haut pour la loi de Fisher décentrée, qui se décompose comme le mélange d'une loi bêta de type II et d'une Poisson.

Exemple 1.2.10. *Considérons $X \sim F_{\nu_1, \nu_2}(\theta)$ pour $\nu_1, \nu_2, \theta > 0$ et posons $f(\cdot; \theta)$ sa densité. En vertu du Lemme 1.1.8 et de l'Exemple 1.2.4, la famille de densités $f(\cdot; \theta)$ a un RVM croissant en X avec paramètre θ et la fonction de masse de $L \stackrel{d}{=} K \mid X \leq t$ a un RVM croissant en L avec paramètre t .*

En effet, pour $\theta' < \theta$ et $x > 0$, on trouve analytiquement que le ratio des densités est

$$\frac{f(x; \theta)}{f(x; \theta')} \propto \frac{\sum_{k \geq 0} (\theta/2)^k \left(\frac{\nu_1 x}{\nu_2 + \nu_1 x} \right)^k \frac{((\nu_1 + \nu_2)/2)_k}{k! (\nu_1/2)_k}}{\sum_{k \geq 0} (\theta'/2)^k \left(\frac{\nu_1 x}{\nu_2 + \nu_1 x} \right)^k \frac{((\nu_1 + \nu_2)/2)_k}{k! (\nu_1/2)_k}},$$

où $(a)_k = \frac{\Gamma(a+k)}{\Gamma(a)}$ est le symbole de Pochhammer¹, est croissant en x . De plus, on a que

$$\frac{\mathbb{P}_\theta(K = k \mid X \leq t)}{\mathbb{P}_{\theta'}(K = k \mid X \leq t)} = \frac{(\theta/2)^k e^{-\theta/2}}{k! \mathbb{P}_\theta(X \leq t)} \bigg/ \frac{(\theta'/2)^k e^{-\theta'/2}}{k! \mathbb{P}_{\theta'}(X \leq t)} \propto (\theta/\theta')^k \quad (1.18)$$

est croissant en k .

1.3 Théorie de la décision statistique

La théorie de la décision statistique est la branche de la statistique qui recherche les stratégies optimales pour prendre une décision afin d'inférer sur un paramètre θ . Dans ce travail, on s'intéresse au problème d'estimation du paramètre θ qui se trouve dans un espace Θ . On note δ , un estimateur appartenant à l'espace des décisions noté D .

1. Aussi appelé factorielle ascendante dans la littérature.

1.3. THÉORIE DE LA DÉCISION STATISTIQUE

Pour déterminer qu'un estimateur est meilleur qu'un autre et quantifier cette amélioration, il est nécessaire d'avoir une fonction dite de perte qu'on note $\rho(\delta, \theta)$. En pratique, on choisit cette fonction principalement pour son comportement : ce qu'elle pénalise et ce qu'elle tolère.

Exemple 1.3.1. *Considérons un échantillon X_1, \dots, X_n i.i.d. f_θ . Alors, pour un vecteur d'observations (x_1, \dots, x_n) , l'estimateur de vraisemblance maximale (EVM) est la solution du problème d'optimisation*

$$\max_{\theta} \sum_{i=1}^n \log f_{\theta}(x_i) \Leftrightarrow \min_{\theta} - \sum_{i=1}^n \frac{1}{n} \log f_{\theta}(x_i)$$

et on reconnaît l'entropie croisée entre f_θ et une uniforme. Cette fonction de perte est particulièrement utilisée pour les tâches de classification en apprentissage automatique comme elle permet de calculer facilement la divergence entre la prédiction et l'information connue a priori. Goodfellow et coll. [10] présentent l'entropie et la perte Kullback-Leibler plus en détail.

Pour quantifier si un estimateur est bon de manière globale et pas seulement ponctuellement, on peut considérer son espérance, ce qui nous indique si δ est en moyenne proche du paramètre qu'il estime. Ainsi, on peut comparer les performances de divers estimateurs entre eux.

Définition 1.3.1. *On dit que $R(\delta, \theta)$ est le risque fréquentiste associé à l'estimateur δ de θ où*

$$R(\delta, \theta) = \mathbb{E}_{\theta}(\rho(\delta, \theta))$$

et $\rho(\delta, \theta)$ est une fonction de perte donnée.

Définition 1.3.2. *Soient δ_1 et δ_2 , deux estimateurs de θ . On dit que δ_1 domine δ_2 si*

$$R(\delta_1, \theta) \leq R(\delta_2, \theta), \quad \forall \theta \in \Theta$$

avec au moins une inégalité stricte pour un $\theta \in \Theta$. Un tel estimateur δ_2 est appelé inadmissible. De plus, un estimateur est dit admissible lorsqu'il n'est pas inadmissible.

1.3. THÉORIE DE LA DÉCISION STATISTIQUE

Quand vient le temps de choisir un «meilleur» estimateur, il est nécessaire de définir un critère d'optimalité. Souvent, on considère l'équivariance ainsi que le critère minimax.

1.3.1 Équivariance

Dans cette section, on considère principalement des pertes invariantes aux changements de position : des fonctions qui prennent seulement en compte l'écart entre l'estimateur $\delta \in \mathbb{R}$ et le paramètre $\theta \in \mathbb{R}$. On les note $\rho(\delta - \theta)$. Toutefois, bien qu'invariantes aux changements de position, il arrive que les fonctions de perte convexes ne soient pas tout à fait adaptées au problème d'estimation. Dans ce cas, on peut se tourner vers une classe plus large : les pertes strictement en bol où $\rho(0) = 0$ puis $\rho(t)$ est croissante pour $t > 0$ et décroissante pour $t < 0$. Cette classe comprend notamment des fonctions de perte bornées telle que la normale réfléchie où

$$\rho(t) = 1 - e^{-t^2/\alpha}, \quad \alpha > 0.$$

On considère pour la suite que $\rho(\delta - \theta)$ est fonction convexe par simplicité. Marchand et Strawderman [57] discutent des technicalités qu'engendre l'utilisation d'une perte strictement un bol. Une présentation plus générale est faite dans le Chapitre 3 du livre de Lehmann et Casella [25] où divers types d'invariance sont présentées pour un échantillon.

Soit $X \in \mathbb{R}^n$ un résumé stochastique de densité

$$f_\theta(x_1, \dots, x_n) = f_0(x_1 - \theta, \dots, x_n - \theta)$$

qu'on dit invariante aux changements de position pour $\theta \in \mathbb{R}$.

Définition 1.3.3. *Pour $X \sim f_0(x_1 - \theta, \dots, x_n - \theta)$, un estimateur δ est dit équivariant par rapport aux changements de position lorsque*

$$\delta(x_1 + c, \dots, x_n + c) = \delta(x_1, \dots, x_n) + c, \quad \forall c, x_1, \dots, x_n \in \mathbb{R}.$$

Exemple 1.3.2. *Parmi d'autres estimateurs équivariants, on retrouve les statistiques*

1.3. THÉORIE DE LA DÉCISION STATISTIQUE

échantillonnables $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$, $\text{Méd}(X_1, \dots, X_n)$ et, plus généralement, les combinaisons convexes de statistiques d'ordre $\sum_{i=1}^n a_i X_{(i)}$ où $\sum_{i=1}^n a_i = 1$ et $a_i \geq 0$ pour $i \in \{1, \dots, n\}$.

Dans cette section, on simplifie avec $n = 1$. Une famille de densités invariante aux changements de position bien connue est la loi normale de variance σ^2 connue. Alors, l'estimateur $\delta(X + c) = X + c$ est dit équivariant : il est invariant aux changements de position et son risque est constant en θ pour des pertes telles que $\rho(\delta - \theta) = |\delta - \theta|^p$ où $p > 0$. Cette dernière propriété est caractéristique des problèmes d'invariance.

Lemme 1.3.3. *Soit $X \sim f_0(x - \theta)$. Alors, pour une fonction de perte $\rho(\delta - \theta)$, on trouve, pour $\delta_c(X) = X + c$, que*

$$R(\delta_c, \theta) = R(\delta_c, 0), \quad \forall \theta \in \Theta.$$

Démonstration. On trouve, en effectuant le changement de variable $z = x - \theta$, que

$$R(\delta_c, \theta) = \int_{\mathbb{R}} \rho(x - c - \theta) f_0(x - \theta) dx = \int_{\mathbb{R}} \rho(z - c) f_0(z) dz = R(\delta_c, 0). \quad \square$$

Définition 1.3.4. *Soient $X \sim f_0(x - \theta)$ avec $f_0 > 0$ et $\rho(\delta - \theta)$ une fonction de perte convexe. On dit que $\delta_*(X) = X - c^*$ est le meilleur estimateur équivariant (MEE) lorsque*

$$R(\delta_*, 0) = \min_{c \in \mathbb{R}} R(\delta_c, 0).$$

Exemple 1.3.4. *Considérons $X \sim f_0(x - \theta)$.*

1. *Pour la perte quadratique, on trouve que le MEE est $X - \mathbb{E}_0(X)$. En effet, on a que*

$$\frac{d}{dc} R(\delta_c, 0) = -2\mathbb{E}_0(X - c) = 0 \Leftrightarrow c = \mathbb{E}_0(X).$$

2. *Pour la perte valeur absolue, on trouve que le MEE est $X - \text{Méd}_0(X)$. En effet, on a que*

$$\frac{d}{dc} R(\delta_c, 0) = \mathbb{P}_0(X > c) - \mathbb{P}_0(X < c) = 0 \Leftrightarrow c = \text{Méd}_0(X).$$

1.3. THÉORIE DE LA DÉCISION STATISTIQUE

À titre d'exemple, lorsque X a pour densité $f_0(x - \theta) = e^{-(x-\theta)} \mathbf{1}_{(\theta, \infty)}(x)$ pour $\theta \in \mathbb{R}$, on trouve que sa médiane est $\log 2$ et on obtient que le MEE est $X - \log 2$.

Typiquement, le MEE satisfait au critère minimax (voir Chapitre 5 de [25] et [31]) qui est présenté à la section ci-dessous. On y retrouvera notamment des estimateurs équivariants dans les exemples donnés.

1.3.2 Règles de Bayes et critère minimax

D'un autre point de vue, on peut considérer le paramètre θ comme une variable aléatoire et lui associer une loi de probabilité qu'on appelle *a priori*.

Définition 1.3.5. *Pour un modèle*

$$\begin{aligned} X \mid \theta &\sim f_\theta \\ \theta &\sim \pi, \end{aligned} \tag{1.19}$$

on appelle π la densité ou la loi *a priori* de θ . Pour une réalisation x , la loi dite *a posteriori* est celle de $\theta \mid X = x$.

De ce modèle, on déduit une classe d'estimateurs : les règles de Bayes. Celles-ci sont intrinsèquement liées à la loi *a priori* ainsi qu'à la fonction de perte. Pour simplifier les calculs, on suppose dans cette section que la convergence dominée de Lebesgue (voir pages 321–322 du livre de Rudin [45]) est satisfaite.

Définition 1.3.6. *Dans le cadre du modèle (1.19), on dit que $\delta_\pi \in D$ est l'estimateur de Bayes associée à la loi *a priori* π pour la perte $\rho(\delta, \theta)$ si*

$$r(\pi, \delta_\pi) = \inf_{\delta \in D} r(\pi, \delta) \tag{1.20}$$

où $r(\pi, \delta) = \mathbb{E}^\theta(R(\delta, \theta))$ est le risque intégré de Bayes. En particulier, lorsque le minimum est unique, on trouve que la règle de Bayes est

$$\delta_\pi(x) = \arg \min_{\delta \in D} \mathbb{E}^\theta(\rho(\delta, \theta) \mid x), \tag{1.21}$$

1.3. THÉORIE DE LA DÉCISION STATISTIQUE

c'est-à-dire qu'elle minimise la perte a posteriori.

En particulier, les règles de Bayes sont généralement uniques lorsqu'on considère des fonctions de perte strictement convexes telles que les écarts quadratique et absolu.

Exemple 1.3.5. Soit $X \in \mathbb{R}$ de densité f_θ avec $\theta \in \mathbb{R}$. Pour la perte valeur absolue et une loi a priori π telle que $\pi(\theta | x) > 0$ sur l'ensemble du domaine, on a que la médiane a posteriori est la règle de Bayes. En effet, on a que

$$\frac{\partial}{\partial \delta} \mathbb{E}^\theta(|\delta - \theta| | X) = 2\mathbb{P}(\theta < \delta | X) - 1 = 0 \Leftrightarrow \delta = \text{Méd}(\theta | X).$$

On considère souvent des fonctions de pertes pondérées afin que le risque puisse être borné. Pour la règle quadratique, la règle de Bayes est la suivante.

Lemme 1.3.6. Pour $X, \theta \in \mathbb{R}$ et la perte quadratique pondérée, lorsque $\rho(\delta, \theta) = \omega(\theta)(\theta - \delta)^2$ avec $\omega(\theta) > 0$, et pour $\theta \sim \pi$, l'estimateur de Bayes δ_π satisfait à

$$\delta_\pi(x) = \frac{\mathbb{E}(\omega(\theta)\theta | X = x)}{\mathbb{E}(\omega(\theta) | X = x)} \quad (1.22)$$

lorsque les espérances $\mathbb{E}(\omega(\theta)\theta | X = x)$, $\mathbb{E}(\omega(\theta) | X = x)$, $\mathbb{E}(\omega(\theta)\theta^2 | X = x)$ existent.

Démonstration. On reprend l'équation (1.21) et on trouve

$$\begin{aligned} \delta_\pi(x) &= \arg \min_{d \in D^*} \mathbb{E}(\omega(\theta)(d - \theta)^2 | X = x) \\ &= \arg \min_{d \in D^*} d^2 \mathbb{E}(\omega(\theta) | X = x) - 2d \mathbb{E}(\omega(\theta)\theta | X = x) \\ &= \frac{\mathbb{E}(\omega(\theta)\theta | X = x)}{\mathbb{E}(\omega(\theta) | X = x)}. \end{aligned} \quad \square$$

Remarque 1.3.7. La règle de Bayes pour la perte quadratique, c'est-à-dire lorsque $\omega(\theta) = 1$, est l'espérance a posteriori.

Un caractère fondamental en théorie de la décision statistique est le critère minimax [49]. Un estimateur minimax et son risque associé témoigne de la difficulté d'estimation liée à un modèle en plus de représenter un point de repère pour l'évaluation de la performance d'autres estimateurs. Lehmann et Casella discutent au

1.3. THÉORIE DE LA DÉCISION STATISTIQUE

Chapitre 5 de [25] du lien entre le critère minimax et l'admissibilité. Comme son nom l'indique, le critère minimax minimise le pire scénario, ce qu'on peut associer, d'un point de vue bayésien, à la loi *a priori* la plus coûteuse au niveau du risque intégré de Bayes $r(\pi, \delta)$.

Définition 1.3.7. *Un estimateur $\delta^* \in D$ est dit minimax lorsque*

$$\sup_{\theta \in \Theta} R(\delta^*, \theta) = \inf_{\delta} \sup_{\theta \in \Theta} R(\delta, \theta).$$

Lorsque la règle de Bayes δ_π est minimax, la loi a priori π (bien qu'elle ne soit pas unique) est dite la moins favorable.

De plus, on a que le critère minimax est fortement lié à la notion de règle de Bayes. En effet, c'est une des principales techniques utilisées pour démontrer qu'un estimateur est minimax en raison des énoncés suivants.

Théorème 1.3.8 (Théorème 1 p. 90 de [8]). *Supposons que δ_π est une règle de Bayes pour $\theta \sim \pi$ et δ telle que $\sup_{\theta} R(\delta, \theta) \leq r(\pi, \delta_\pi) < \infty$, alors δ est minimax et π est la loi a priori la moins favorable.*

Démonstration. On procède par contradiction. Supposons que δ ne soit pas minimax. Alors il existe une règle $\tilde{\delta}$ telle que $\sup_{\theta \in \Theta} R(\tilde{\delta}, \theta) < \sup_{\theta \in \Theta} R(\delta, \theta)$. Or, comme le risque intégré de Bayes est une espérance, on a que $r(\pi, \tilde{\delta}) \leq \sup_{\theta \in \Theta} R(\tilde{\delta}, \theta)$. D'où,

$$r(\pi, \tilde{\delta}) \leq \sup_{\theta \in \Theta} R(\tilde{\delta}, \theta) < \sup_{\theta} R(\delta, \theta) \leq r(\pi, \delta_\pi).$$

Or, δ_π est une règle de Bayes si et seulement si $r(\pi, \delta_\pi) = \inf_d r(\pi, d)$ et on obtient une contradiction. Ainsi, δ est minimax. \square

Corollaire 1.3.9. *Supposons δ_π un estimateur de Bayes pour $\theta \sim \pi$. Lorsqu'il existe $c \in \mathbb{R}$ tel que $R(\delta_\pi, \theta) = c$ pour tout θ , alors δ_π est minimax.*

Le Corollaire 1.3.9 est le plus facile à appliquer pour démontrer qu'un estimateur est minimax. Toutefois, trouver des règles de Bayes qui ont des risques constants peut nécessiter quelques manipulations telles que la pondération de la fonction de perte et le changement de loi *a priori*.

1.3. THÉORIE DE LA DÉCISION STATISTIQUE

Lemme 1.3.10. Soient $X \mid \theta \sim f_\theta$ avec $X, \theta \in \mathbb{R}$, les densités a priori $\pi_1(\theta)$ et $\pi_0(\theta) \propto \omega(\theta)\pi_1(\theta)$ avec $\omega(\theta) > 0$, et les fonctions de perte $\rho_0(\delta, \theta)$ et $\rho_1(\delta, \theta) = \omega(\theta)\rho_0(\delta, \theta)$. Si les règles de Bayes δ_{π_i} respectivement associées aux fonctions de pertes ρ_i ($i = 0, 1$), sont finies et uniques, alors $\delta_{\pi_0} = \delta_{\pi_1}$. De plus, on a que $r(\pi_0, \delta_{\pi_0}) = \frac{r(\pi_1, \delta_{\pi_1})}{\int_{\Theta} \pi_1(\theta)\omega(\theta)d\theta}$ lorsque ces quantités convergent.

Démonstration. On a que les densités a posteriori sont $\pi_1(\theta \mid X = x) = \frac{f_\theta(x)\pi_1(\theta)}{\int_{\Theta} f_\theta(x)\pi_1(\theta)d\theta}$ et $\pi_0(\theta \mid X = x) = \frac{f_\theta(x)\omega(\theta)\pi_1(\theta)}{\int_{\Theta} f_\theta(x)\omega(\theta)\pi_1(\theta)d\theta}$. On trouve donc que

$$\begin{aligned} \delta_{\pi_1}(x) &= \arg \min_{\delta} \mathbb{E}(\rho_1(\delta, \theta) \mid X = x) \\ &= \arg \min_{\delta} \frac{1}{\int_{\Theta} f_\theta(x)\pi_1(\theta)d\theta} \int_{\Theta} \omega(\theta)\rho_0(\delta, \theta)f_\theta(x)\pi_1(\theta)d\theta \\ &= \arg \min_{\delta} \frac{1}{\int_{\Theta} f_\theta(x)\pi_0(\theta)d\theta} \int_{\Theta} \rho_0(\delta, \theta)f_\theta(x)\pi_0(\theta)d\theta \\ &= \arg \min_{\delta} \mathbb{E}(\rho_0(\delta, \theta) \mid X = x) \\ &= \delta_{\pi_0}(x). \end{aligned}$$

De plus, on a que le risque intégré de Bayes de δ_{π_1} est

$$\begin{aligned} r(\pi_1, \delta_{\pi_1}) &= \int_{\Theta} \pi_1(\theta) \int_X \rho_1(\delta_{\pi_1}, \theta) f_\theta(x) dx d\theta \\ &= \int_{\Theta} \omega(\theta)\pi_1(\theta) \int_X \rho_0(\delta_{\pi_1}, \theta) f_\theta(x) dx d\theta \\ &= \int_{\Theta} \omega(\theta)\pi_1(\theta) d\theta \int_{\Theta} \pi_0(\theta) \int_X \rho_0(\delta_{\pi_1}, \theta) f_\theta(x) dx d\theta \\ &= r(\pi_0, \delta_{\pi_0}) \int_{\Theta} \omega(\theta)\pi_1(\theta) d\theta. \end{aligned} \quad \square$$

Exemple 1.3.11. Considérons $X \mid \theta \sim \mathcal{P}(\theta)$ avec $\theta \sim \mathcal{G}(\alpha, \beta)$ où $\alpha, \beta > 0$. On trouve que la densité a posteriori est

$$\pi_0(\theta \mid X = x) \propto \frac{e^{-\theta} \theta^x}{x!} \frac{\theta^{\alpha-1} e^{-\frac{\theta}{\beta}}}{\Gamma(\alpha)\beta^\alpha} \mathbf{1}_{(0, \infty)}(\theta) \propto \theta^{\alpha+x-1} e^{-\theta(\frac{\beta+1}{\beta})} \mathbf{1}_{(0, \infty)}(\theta),$$

et on reconnaît le noyau d'une $\mathcal{G}(x + \alpha, \beta/(\beta + 1))$. Ainsi, $\mathbb{E}(\theta \mid X = x) = \frac{\beta}{\beta+1}x + \frac{\alpha\beta}{\beta+1}$ est une règle de Bayes sous la perte quadratique. De plus, son risque fréquentiste est

1.3. THÉORIE DE LA DÉCISION STATISTIQUE

$R(\delta_{\pi_0}, \theta) = \left(\frac{\alpha\beta}{\beta+1}\right)^2 + \left(\frac{\beta}{\beta+1}\right)^2 \theta + \frac{1-2\beta}{(\beta+1)^2} \theta^2$ et on déduit que son risque intégré de Bayes est $r(\pi_0, \delta_{\pi_0}) = \left(\frac{\alpha\beta}{\beta+1}\right)^2 + \left(\frac{\beta}{\beta+1}\right)^2 \alpha\beta + \frac{1-2\beta}{(\beta+1)^2} \beta^2 \alpha(\alpha+1)$.

On a que δ_{π_0} est aussi la règle de Bayes pour la perte quadratique normalisée $\rho_1(\delta - \theta) = \frac{(\delta - \theta)^2}{\theta}$ lorsqu'on considère $\pi_1(\theta) = \frac{\theta^\alpha e^{-\theta/\beta}}{\Gamma(\alpha+1)\beta^{\alpha+1}} \mathbb{1}_{(0,\infty)}(\theta)$. En effet, on trouve que $\theta \mid X = x \sim \mathcal{G}(\alpha + x + 1, \beta/(\beta + 1))$ et que

$$\delta_{\pi_1}(x) = \frac{1}{\mathbb{E}(1/\theta \mid X = x)} = \frac{1}{\frac{\beta+1}{\beta}/(x + \alpha)} = \delta_{\pi_0}(x). \quad (1.23)$$

De plus, son risque intégré de Bayes est

$$r(\pi_1, \delta_{\pi_1}) = \frac{1}{\alpha\beta} \left(\frac{\alpha\beta}{\beta+1}\right)^2 + \left(\frac{\beta}{\beta+1}\right)^2 + \frac{1-2\beta}{(\beta+1)^2} \beta(\alpha+1) = \frac{1}{\alpha\beta} r(\pi_0, \delta_{\pi_0}). \quad (1.24)$$

Pour les règles de Bayes, on a besoin d'avoir recours à la généralisation suivante.

Définition 1.3.8. Considérons le modèle (1.19). On dit que π est une loi a priori impropre lorsque $\int_{\Theta} \pi(\theta) d\theta = \infty$. Dans ce cas, la règle de Bayes associée à π définie par l'équation (1.21) est dite généralisée.

Exemple 1.3.12. On s'intéresse aux lois uniformes sur un intervalle non borné ou semi-borné : des distributions a priori impropres parmi les plus simples.

1. Considérons $X \mid \theta \sim f_0(x - \theta)$ avec $\pi(\theta) = 1$. Pour ce modèle, on trouve que la loi a posteriori est

$$\pi(\theta \mid X = x) \propto f_0(x - \theta).$$

Pour la perte $\rho(\delta - \theta) = (\delta - \theta)^2$, on trouve que la règle de Bayes généralisée est

$$\delta_{\pi}(x) = \int \theta \frac{f_0(x - \theta)}{\int f_0(x - \theta) d\theta} d\theta = \int (x - z) f_0(z) dz = x - \mathbb{E}_0(X)$$

et on reconnaît le MEE de l'Exemple 1.3.4.1. Une présentation plus complète des règles de Bayes équivariantes est donnée à la section 4.4 de [25].

2. Considérons $X \mid \theta \sim \mathcal{P}(\theta)$ avec $\pi(\theta) = \mathbb{1}_{[\theta_0, \infty)}(\theta)$ pour $\theta_0 \geq 0$. Pour ce modèle, on trouve que la loi a posteriori est

$$\pi(\theta \mid X = x) \propto \frac{\theta^x e^{-\theta}}{x!} \mathbb{1}_{[\theta_0, \infty)}(\theta) \propto \theta^x e^{-\theta} \mathbb{1}_{[\theta_0, \infty)}(\theta).$$

1.3. THÉORIE DE LA DÉCISION STATISTIQUE

Pour la perte $\rho(\delta - \theta) = \frac{(\delta - \theta)^2}{\theta}$, on trouve, en intégrant par parties puis en effectuant le changement de variables $t = \frac{\theta}{\theta_0}$, que la règle de Bayes généralisée est

$$\begin{aligned}\delta_\pi(x) &= \frac{1}{\mathbb{E}(1/\theta \mid X = x)} = \frac{1}{\int_{\theta_0}^{\infty} \theta^{x-1} e^{-\theta} d\theta / \int_{\theta_0}^{\infty} \theta^x e^{-\theta} d\theta} \\ &= x + \frac{1}{\int_1^{\infty} t^{x-1} e^{-\theta_0(t-1)} dt}.\end{aligned}$$

Parmi les lois *a priori* impropres généralement utilisées, on retrouve fréquemment la loi de Jeffreys, une des lois dites non informatives. Carlin et Louis [4] discutent de ce type de lois plus en détail à la section 2.2. Pour la loi de Poisson présentée à l'Exemple 1.3.11, on considérera plutôt à l'Exemple 1.3.14 une suite de lois *a priori* $(\pi_n)_{n \geq 0}$ qui converge asymptotiquement vers une loi impropre.

Théorème 1.3.13. *Soient δ un estimateur de θ et $(\pi_n)_{n \geq 0}$ une suite de lois *a priori* associées aux règles de Bayes δ_{π_n} . Si $\sup_{\theta} R(\delta, \theta) = \lim_{n \rightarrow \infty} r(\pi_n, \delta_{\pi_n})$, alors δ est minimax et $(\pi_n)_{n \geq 0}$ est la suite de lois *a priori* la moins favorable.*

Exemple 1.3.14 (Suite de l'Exemple 1.3.11). *Lorsque $X \mid \theta \sim \mathcal{P}(\theta)$ avec $\theta \geq 0$, on montre que X est minimax pour la fonction de perte normalisée $\rho(\delta - \theta) = \frac{(\delta - \theta)^2}{\theta}$.*

*En vertu de l'équation (1.23), on obtient que X est une limite de règle de Bayes lorsque $\alpha \rightarrow 0$ et $\beta \rightarrow \infty$. Considérons alors la suite de lois *a priori* $\pi_n \sim \mathcal{G}(1+1/n, n)$ pour $n > 0$. Lorsqu'on reprend l'équation (1.24), on trouve que*

$$r(\pi_n, \delta_{\pi_n}) = \left(\frac{1}{n+1}\right)^2 + \left(\frac{n}{n+1}\right)^2 + \frac{1-2n}{(n+1)^2} n(1+1/n) \rightarrow 1, \quad \text{lorsque } n \rightarrow \infty.$$

Puisque

$$R(X, \theta) = \mathbb{E}_{\theta} \left[\frac{(X - \theta)^2}{\theta} \right] = \frac{\text{Var}_{\theta}(X)}{\theta} = 1,$$

X est une règle de Bayes par extension à risque constant et est minimax pour la perte pondérée $\rho(\delta - \theta) = \frac{(\delta - \theta)^2}{\theta}$ en vertu du Théorème 1.3.13. Cette technique est malheureusement plus difficile à utiliser pour démontrer que X demeure minimax pour cette perte en présence d'une borne inférieure de l'espace paramétrique.

1.3. THÉORIE DE LA DÉCISION STATISTIQUE

Une autre technique pour déterminer qu'un estimateur est minimax requiert la borne de Cramér-Rao.

Théorème 1.3.15 (Lemme 5.3 de [25]). *Soit $X \sim f_\theta$ tel que $f_\theta > 0$, $\frac{d}{d\theta} \log f_\theta$ existe et est fini et $\frac{d}{d\theta} \int f_\theta(x) dx = \int \frac{d}{d\theta} f_\theta(x) dx$. Alors, pour un estimateur δ de biais $b(\theta) = \mathbb{E}_\theta(\delta) - \theta \in \mathbb{R}$, on a*

$$\text{Var}_\theta(\delta) \geq \frac{(1 + b'(\theta))^2}{\mathcal{I}(\theta)}, \quad (1.25)$$

où $\mathcal{I}(\theta) = -\mathbb{E}_\theta \left[\frac{\partial^2}{\partial \theta^2} \log f_\theta(X) \right]$.

Exemple 1.3.16 (Théorème 4.1 de [51], suite de l'Exemple 1.3.14). *Lorsque $\theta \geq \theta_0 > 0$, on peut montrer que X demeure minimax en travaillant avec la borne de Cramér-Rao. On reprend la démonstration de [51] aux pages 57–58 en procédant par contradiction.*

Supposons qu'il existe $\epsilon \in (0, 1)$ et un estimateur δ' , de biais $b(\theta)$, tels que, pour tout $\theta \geq \theta_0$,

$$R(\delta', \theta) = \mathbb{E}_\theta \left(\frac{(\delta' - \theta)^2}{\theta} \right) \leq 1 - \epsilon < R(X, \theta) = 1.$$

Comme $\mathcal{I}(\theta) = \frac{1}{\theta}$ lorsque $X \sim \mathcal{P}(\theta)$, on a que

$$R(\delta', \theta) = \frac{\text{Var}_\theta(\delta') + b^2(\theta)}{\theta} \geq (1 + b'(\theta))^2 + \frac{b'(\theta)}{\theta}, \quad \forall \theta \geq \theta_0$$

en vertu de l'inégalité (1.25). On trouve donc, pour tout $\theta \geq \theta_0$, que

$$-\sqrt{1 - \epsilon} - 1 \leq b'(\theta) \leq \sqrt{1 - \epsilon} - 1 \quad (1.26)$$

$$-\sqrt{\theta(1 - \epsilon)} \leq b(\theta) \leq \sqrt{\theta(1 - \epsilon)}. \quad (1.27)$$

On peut déduire de (1.27) que $0 \leq b(\theta) + \sqrt{\theta(1 - \epsilon)}$. Si on dérive cette quantité, on trouve, pour tout $\theta \geq \theta_0$ que

$$\frac{d}{d\theta} \left(b(\theta) + \sqrt{\theta(1 - \epsilon)} \right) = b'(\theta) + \frac{\sqrt{1 - \epsilon}}{2\sqrt{\theta}} \leq \sqrt{1 - \epsilon} - 1 + \frac{\sqrt{1 - \epsilon}}{2\sqrt{\theta}}$$

en vertu de (1.26).

1.4. MÉTHODE DE KUBOKAWA

Posons $\theta^* > \max \left\{ \theta_0, \frac{1-\epsilon}{(1-\sqrt{1-\epsilon})^2} \right\}$. Ainsi, on obtient que $\frac{1}{\sqrt{\theta}} < \min \left(\frac{1}{\sqrt{\theta_0}}, \frac{1-\sqrt{1-\epsilon}}{\sqrt{1-\epsilon}} \right)$ pour tout $\theta \geq \theta^*$, d'où,

$$\frac{d}{d\theta} (b(\theta) + \sqrt{\theta(1-\epsilon)}) < \sqrt{1-\epsilon} - 1 + \frac{1 - \sqrt{1-\epsilon}}{2} = -\frac{1 - \sqrt{1-\epsilon}}{2} < 0.$$

Comme il est possible de borner supérieurement la dérivée de $b(\theta) + \sqrt{\theta(1-\epsilon)}$ par une constante négative pour tout $\theta \geq \theta_0$, il est impossible de satisfaire simultanément la condition $0 \leq b(\theta) + \sqrt{\theta(1-\epsilon)}$. On a donc, par contradiction, que X demeure minimax.

De plus, comme l'EVM $\lambda_{EVM} = \max\{\theta_0, X\}$ domine X , ce dernier est aussi minimax sous la perte quadratique normalisée $\frac{(\delta-\theta)^2}{\theta}$.

1.4 Méthode de Kubokawa

Les développements présentés dans cette section reposent sur une adaptation de la méthode de Kubokawa [21]. Cette dernière permet de déduire des conditions de dominance pour un estimateur δ_0 donné. L'élément clé de cette technique est l'expression d'une différence de fonctions de perte comme une intégrale, ce qui permet d'exprimer la différence des risques fréquentistes de deux estimateurs comme une intégrale double. Cette méthode est connue sous l'acronyme IERD (*Integral Expression for Risk Difference*). L'article de revue de Kubokawa [22] offre une bonne vue d'ensemble des multiples applications de cette méthode dont l'estimation de paramètres de position ou d'échelle sous contraintes, l'estimation de Stein, l'estimation de paramètres de décentralité, l'estimation dans des modèles de lois discrètes, l'estimation par intervalles de confiance et l'estimation de densités prédictives [24]. Comme elle repose sur peu d'hypothèses et est flexible quant au choix de la fonction de perte, cette technique a pu aussi être utilisée pour la démonstration de la non minimaxité d'un estimateur donné (pp. 58–59 du livre de van Eeden [51]) et la construction d'estimateurs minimax pour des paramètres ayant une borne inférieure (voir les articles de Marchand et Strawderman [56] section 3.1 et [57]), pour ne citer que ceux-ci.

Dans cette section, on illustre la méthode de Kubokawa pour l'estimation d'un paramètre borné inférieurement. Une première application fait intervenir un modèle

1.4. MÉTHODE DE KUBOKAWA

de loi continue, tandis que la seconde se rapporte à un modèle de loi discrète et notamment l'estimation de la moyenne d'une loi de Poisson bornée inférieurement telle qu'introduite à la fin de la section précédente.

1.4.1 Pour les lois continues

Pour utiliser cette technique, il est nécessaire que les fonctions considérées soient absolument continues. En effet, l'astuce sur laquelle repose la méthode de Kubokawa est en réalité une application assez directe de cette définition.

Définition 1.4.1. *On dit que f est une fonction absolument continue sur $I \subset \mathbb{R}$ si et seulement s'il existe une fonction g intégrable sur I telle que, pour tout $a < x \in I$,*

$$f(x) - f(a) = \int_a^x g(t)dt.$$

Autrement dit, sous la mesure de Lebesgue, $f' = g$ presque partout.

Remarque 1.4.1. *On utilise f' par la suite pour alléger la notation.*

Lemme 1.4.2 (IERD de Kubokawa). *Soient ρ et ψ deux fonctions absolument continues sur \mathbb{R} et $t \in \mathbb{R}$. Supposons que $\lim_{w \rightarrow \infty} \psi(w) = 0$. Alors, pour δ_0 un estimateur de θ , la différence des pertes est donnée par*

$$\rho(\delta_0(X) - \theta) - \rho(\delta_0(X) - \psi(t) - \theta) = - \int_t^\infty \rho'(\delta_0(X) - \psi(w) - \theta) \psi'(w) dw.$$

Démonstration. En utilisant w comme variable d'intégration, on trouve que

$$\begin{aligned} \rho(\delta_0(X) - \theta) - \rho(\delta_0(X) - \psi(t) - \theta) &= \rho(\delta_0(X) - \psi(w) - \theta) \Big|_t^\infty \\ &= - \int_t^\infty \rho'(\delta_0(X) - \psi(w) - \theta) \psi'(w) dw. \quad \square \end{aligned}$$

On illustre ici la méthode de Kubokawa lorsque $\theta \geq \theta_0$ pour un θ_0 connu et $X \sim f_0(x - \theta)$ une famille de densités de position. Pour estimer θ , on considère le meilleur estimateur équivariant (MEE) $\delta_0(X) = X - c$ pour une constante $c \in \mathbb{R}$. Cette dernière sera ajustée par rapport à la fonction de perte tel que développé dans la sous-section [1.3.1](#).

1.4. MÉTHODE DE KUBOKAWA

On recherche des estimateurs $\delta_\psi(X) = X - \psi(X)$ tels qu'ils dominent $X - c$ pour une fonction de perte $\rho(\delta - \theta)$. Posons $C = \{\delta_\psi : \delta_\psi(X) = X - \psi(X)\}$, une classe d'estimateurs de θ .

On s'intéresse ici au cas où :

- (a) la fonction de perte $\rho(\delta - \theta)$ est absolument continue et convexe ;
- (b) $\delta_\psi \in C$ où ψ est une fonction strictement croissante et absolument continue telle que $\lim_{t \rightarrow \infty} \psi(t) = c$.

Les cas où la perte n'est pas nécessairement convexe et où l'on est en présence d'un échantillon avec densité $f_0(x_1 - \theta, \dots, x_n - \theta)$ sont traités dans l'article de Marchand et Strawderman [57]. Dans ce contexte, on ne peut autant réduire les expressions sans ajouter de contraintes supplémentaires. En effet, un argument est nécessaire pour assurer que le *supremum* sur θ de l'espérance conditionnelle est atteint à la borne inférieure de l'espace paramétrique.

Lemme 1.4.3. *Soient $X \sim f_0(x - \theta)$ pour $\theta \geq \theta_0$ et $\rho(\delta - \theta)$ satisfaisant à la condition (a). Soit $\delta_0(X) = X - c$ le MEE de risque $R(\delta_0, 0) = R$. Alors, on a que $\delta_\psi = X - \psi(X)$ vérifiant la condition (b) domine $\delta_0(X)$ lorsque*

$$\mathbb{E}_0 [\rho'(X - \psi(w)) \mid X \leq w - \theta_0] \leq 0, \quad \forall w \quad (1.28)$$

si la condition

$$\int_{-\infty}^{\infty} \int_t^{\infty} |\rho'(t - \psi(w) - \theta)| \psi'(w) f_0(t - \theta) dw dt < \infty \quad (1.29)$$

est satisfaite.

Remarque 1.4.4. *On a que le théorème de Fubini est vérifié avec l'inégalité (1.29), on peut donc interchanger les intégrales. Dans l'article de Kubokawa [21], on suppose qu'il existe $\alpha > 0$ tel que*

- $\mathbb{E}(\rho(\alpha X - \theta)) < \infty$;
- $\mathbb{E}(|\rho'(\alpha X - \theta)|) < \infty$.

1.4. MÉTHODE DE KUBOKAWA

Démonstration du Lemme 1.4.3. En vertu du Lemme 1.4.2, on trouve que

$$\begin{aligned} R(\delta_0, \theta) - R(\delta_\psi, \theta) &= \mathbb{E}_\theta [\rho(X - c - \theta) - \rho(X - \psi(X) - \theta)] \\ &= - \int_{-\infty}^{\infty} \left[\int_t^{\infty} \rho'(t - \psi(w) - \theta) \psi'(w) dw \right] f_0(t - \theta) dt \\ &= - \int_{-\infty}^{\infty} \left[\int_{-\infty}^w \rho'(t - \psi(w) - \theta) f_0(t - \theta) dt \right] \psi'(w) dw. \end{aligned}$$

On a alors que $\delta_\psi(X)$ domine $X - c$ lorsque $R(\delta_0, \theta) - R(\delta_\psi, \theta) \geq 0$, soit si et seulement si

$$\int_{-\infty}^{\infty} \left[\int_{-\infty}^w \rho'(t - \psi(w) - \theta) f_0(t - \theta) dt \right] \psi'(w) dw \leq 0, \quad \forall \theta \geq \theta_0. \quad (1.30)$$

Comme $\psi(w)$ est une fonction non décroissante, une condition suffisante pour que l'inégalité (1.30) soit vérifiée est que

$$\begin{aligned} \int_{-\infty}^w \rho'(t - \psi(w) - \theta) f_0(t - \theta) dt &\leq 0, \quad \forall w, \theta \geq \theta_0 \\ \Leftrightarrow \int_{-\infty}^{w-\theta} \rho'(t - \psi(w)) f_0(t) dt &\leq 0, \quad \forall w, \theta \geq \theta_0 \\ \Leftrightarrow \sup_{\theta \geq \theta_0} \mathbb{E}_0 [\rho'(X - \psi(w)) \mid X \leq w - \theta] &\leq 0, \quad \forall w. \end{aligned}$$

À l'aide du Lemme 1.2.6, on trouve que la densité de $Y \stackrel{d}{=} X \mid X \leq w - \theta$ a un RVM décroissant en Y avec paramètre θ . Puisque ρ est convexe, sa dérivée est croissante et l'espérance conditionnelle $\mathbb{E}_0 [\rho'(X - \psi(w)) \mid X \leq w - \theta]$ est décroissante en θ en vertu des lemmes 1.2.5 et 1.2.2. On a donc que δ_ψ domine le MEE $X - c$ si

$$0 \geq \sup_{\theta \geq \theta_0} \mathbb{E}_0 [\rho'(X - \psi(w)) \mid X \leq w - \theta] = \mathbb{E}_0 [\rho'(X - \psi(w)) \mid X \leq w - \theta_0]. \quad \square$$

Comme Marchand et Strawderman [57] ont démontré que le MEE est minimax pour des familles de densités invariantes aux changements de position sur des espaces euclidiens en présence de la contrainte $\theta \geq \theta_0$, on obtient que δ_ψ l'est aussi. On cherche maintenant à déterminer si δ_ψ est un estimateur bayésien.

Posons $\delta_U(X) = X - \psi_U(X)$ la règle de Bayes généralisée lorsque la loi *a priori* est une uniforme sur $[\theta_0, \infty)$. Pour trouver une caractérisation de δ_U , il est nécessaire de permuter la dérivée et l'intégrale, ce qui peut notamment être assuré par la

1.4. MÉTHODE DE KUBOKAWA

convergence dominée de Lebesgue (voir pages 321–322 de [45]).

Lemme 1.4.5. *Supposons $X \mid \theta \sim f_0(x - \theta)$. Pour une fonction de perte ρ satisfaisant à la condition (a) et, pour un x observé, $\delta_U(x) = x - \psi_U(x)$ la règle de Bayes généralisée associée à l'a priori $\pi(\theta) = \mathbb{1}_{\theta_0, \infty}(\theta)$, on a que*

$$\mathbb{E}_0(\rho'(X - \psi_U(x)) \mid X \leq x - \theta_0) = 0, \quad \forall x \quad (1.31)$$

lorsqu'il existe $c \in \mathbb{R}$ tel que $\max\{\mathbb{E}_0(\rho(X - c)), \mathbb{E}_0(\rho'(X - c))\} < \infty$.

Démonstration. Soit $X - \psi(x)$, un estimateur de θ . On veut minimiser, pour tout x , l'espérance *a posteriori* $\mathbb{E}_{\theta_0}(\rho(X - \psi) \mid X \leq x - \theta_0)$ en ψ . Comme la convergence dominée est vérifiée, on dérive à l'intérieur de l'intégrale pour obtenir

$$\begin{aligned} \frac{d}{d\psi} \mathbb{E}_0(\rho(X - \psi) \mid X \leq x - \theta_0) &= \frac{d}{d\psi} \frac{\int_{-\infty}^{x-\theta_0} \rho(t - \psi) f_0(t - \theta_0) dt}{\mathbb{P}_0(X \leq x - \theta_0)} \\ &= -\mathbb{E}_0(\rho'(X - \psi) \mid X \leq x - \theta_0). \end{aligned} \quad (1.32)$$

Comme ρ est une fonction convexe, le terme de droite de l'équation (1.32) est une fonction croissante en ψ . Par définition, on a que la règle de Bayes minimise l'espérance de la perte *a posteriori*, d'où

$$\mathbb{E}_0(\rho'(X - \psi_U(x)) \mid X \leq x - \theta_0) = 0. \quad \square$$

On a précédemment vu à la démonstration du Lemme 1.4.3 qu'une condition suffisante pour que l'estimateur δ_ψ domine $X - c$ est que

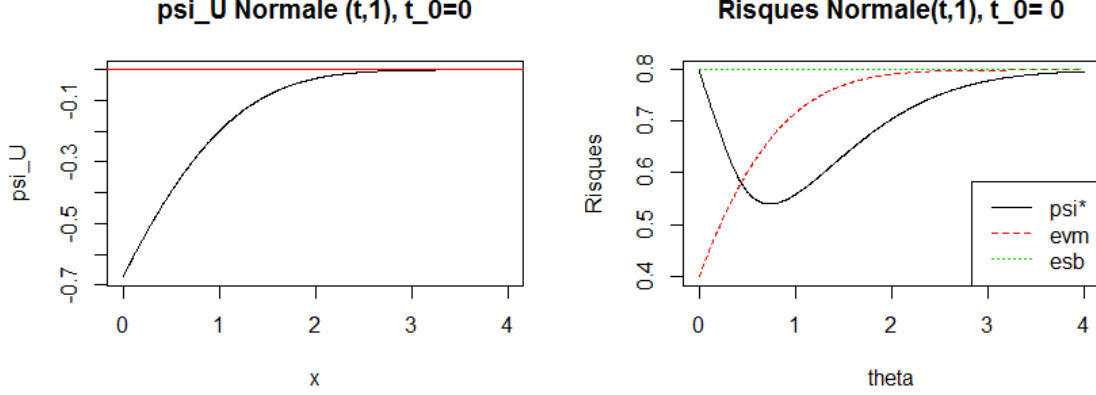
$$\mathbb{E}_0[\rho'(X - \psi(w)) \mid X \leq w - \theta_0] \leq 0, \quad \forall w, \theta \geq \theta_0.$$

Ainsi, la règle de Bayes représente la frontière de la classe d'estimateurs C .

Corollaire 1.4.6. *Dans le contexte du Lemme 1.4.5, la règle de Bayes $\delta_U(X) = X - \psi_U(X)$ domine $X - c$ pour une perte $\rho(\delta - \theta)$ satisfaisant à la condition (a).*

Il est important de noter que dans le cas d'une perte non convexe, il est théoriquement possible d'avoir plusieurs règles de Bayes. Toutefois, lorsque ρ est strictement

1.4. MÉTHODE DE KUBOKAWA



(a) Fonction ψ_U , la borne inférieure de ψ .

(b) Risques.

Figure 1.1 – Fonction ψ_U et risques d’une normale de moyenne θ et d’écart-type 1 lorsque $\theta \geq 0$ pour la perte valeur absolue.

convexe, δ_U est unique. L’unicité des règles de Bayes pour d’autres fonctions de perte est abordée dans [57].

Exemple 1.4.7. *Pour illustrer les résultats ci-dessus, considérons un échantillon X_1, \dots, X_n i.i.d $\mathcal{N}(\theta, \sigma^2)$ avec $\sigma > 0$ connue et $\theta \geq 0$. Alors, $\bar{X}_n \stackrel{d}{=} \frac{1}{n} \sum_{i=1}^n X_i \sim \mathcal{N}(\theta, \sigma^2/n)$. On souhaite estimer θ à l’aide du résumé exhaustif \bar{X}_n .*

Pour la perte valeur absolue, on souhaite construire un estimateur $\delta_\psi \in C$ dominant le MEE $\delta_0(\bar{X}_n) = \bar{X}_n - \text{Méd}_0(\bar{X}_n) = \bar{X}_n$. Comme $\rho'(t) = \text{sgn}(t) = \mathbb{1}_{(0,\infty)}(x) - \mathbb{1}_{(-\infty,0)}(x)$ presque partout, on a, en appliquant le Lemme 1.4.3, que δ_ψ domine \bar{X}_n lorsque

$$\begin{aligned} \mathbb{E}_0 \left[\text{sgn}(\bar{X}_n - \psi(w)) \mid \bar{X}_n \leq w \right] &= 2\mathbb{P}_0 \left[\bar{X}_n > \psi(w) \mid \bar{X}_n \leq w \right] - 1 \leq 0, \quad \forall w \in \mathbb{R} \\ \Leftrightarrow \psi(w) &\geq \text{Méd}_0(\bar{X}_n \mid \bar{X}_n \leq w) = \Phi^{-1} \left(\frac{\Phi(w\sqrt{n}/\sigma)}{2} \right), \quad \forall w \in \mathbb{R} \end{aligned}$$

où Φ est la fonction de répartition d’une loi normale centrée et réduite. On a bien que $\lim_{w \rightarrow \infty} \Phi^{-1} \left(\frac{\Phi(w\sqrt{n}/\sigma)}{2} \right) = 0$ et que la médiane tronquée est une fonction non décroissante, comme on peut le remarquer à la figure 1.1a.

Les risques de δ_U , δ_0 et $\delta_{EVM}(Y) = \max\{\theta_0, Y\}$ sont illustrés à la figure 1.1b

1.4. MÉTHODE DE KUBOKAWA

lorsque $\sigma = \sqrt{n}$. On remarque que le gain maximal de δ_U est assez rapproché de la borne inférieure de l'espace paramétrique, lorsque $\theta \approx 0,731$, où l'on a une amélioration supérieure à 30% et que δ_{EVM} performe mieux que Y sur un petit intervalle pour de faibles valeurs de θ , entre 0 et 0,438. On observera ce schéma pour tous les graphiques de risques obtenus en appliquant cette méthode. De plus, on note que le rapport est égal à 1 lorsque $\theta = 0$ et $\theta \rightarrow \infty$.

Pour la plupart des fonctions de perte, il est difficile d'isoler le terme ψ_U de l'espérance. On a remarqué à l'Exemple 1.4.7 que ce n'est pas une difficulté lorsqu'on considère la perte valeur absolue.

Théorème 1.4.8. *Pour la perte valeur absolue et $X \sim f_0(x - \theta)$ avec $f_0 > 0$ et $\theta \geq \theta_0$, l'estimateur $\delta_\psi(X) = X - \psi(X)$ satisfaisant la condition (b) domine le MEE $X - c$ lorsque*

$$F_0^{-1} \left(\frac{F_0(x - \theta_0)}{2} \right) + \theta_0 \leq \psi(x), \quad \forall x,$$

où F_0 est la fonction de répartition associée à f_0 .

Démonstration. On reprend l'équation (1.31) et on trouve, pour tout x ,

$$\begin{aligned} 0 &= \mathbb{E}_0(\text{sgn}(X - \psi_U(x)) \mid X \leq x - \theta_0) \\ &= \int_{-\infty}^{\psi_U(x)} \frac{f_0(z - \theta_0)}{F_0(x - \theta_0)} dz - \int_{\psi_U(x)}^{x - \theta_0} \frac{f_0(z - \theta_0)}{F_0(x - \theta_0)} dz \\ &= 1 - 2\mathbb{P}_0(\psi_U(x) \leq X \leq x - \theta_0 \mid X \leq x - \theta_0). \end{aligned}$$

Comme $f_0 > 0$, la fonction de répartition F_0 est inversible et on trouve que

$$\psi_U(x) = \text{Méd}_0(X \mid X \leq x - \theta_0) = F_0^{-1} \left(\frac{F_0(x - \theta_0)}{2} \right) + \theta_0$$

où $\text{Méd}(\cdot)$ est la médiane. En isolant ψ dans l'inégalité (1.28), on a que

$$\psi(x) \geq F_0^{-1} \left(\frac{F_0(x - \theta_0)}{2} \right) + \theta_0 = \psi_U(x), \quad \forall x$$

est une condition suffisante pour que δ_ψ domine δ_0 . □

1.4. MÉTHODE DE KUBOKAWA

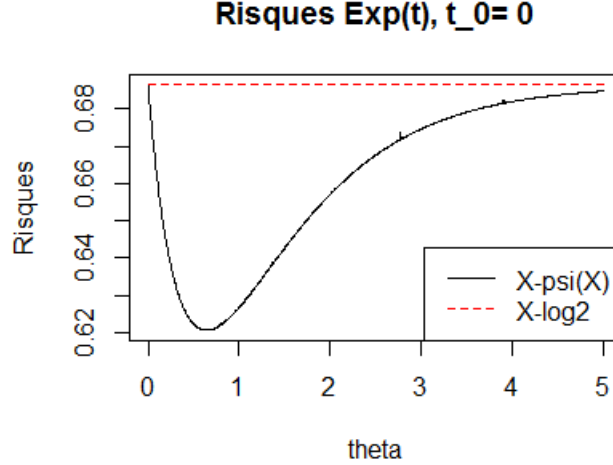


Figure 1.2 – Risques pour la perte valeur absolue dans le cas d’une exponentielle de position de paramètre $\theta > 0$.

Dans le prochain exemple, on s’intéresse à la famille de position d’une loi exponentielle.

Exemple 1.4.9. *Considérons un échantillon X_1, \dots, X_n i.i.d. de densité $f_\theta(x) = f_0(x - \theta) = \beta e^{-\beta(x-\theta)} \mathbf{1}_{(0, \infty)}(x - \theta)$. On note $X_i \sim \text{Exp}(\beta, \theta)$. Alors, on trouve que $Y \stackrel{d}{=} n\beta \min_{1 \leq i \leq n} X_i \sim \text{Exp}(1, \theta)$. On a démontré à l’Exemple 1.3.4.2 que le MEE est $\delta_0(Y) = Y - \log 2$. On cherche donc à dominer ce dernier plutôt que simplement Y . À l’aide du Théorème 1.4.8, on obtient la condition suffisante*

$$\psi(w) \geq \text{Méd}_0(Y \mid Y \leq w) = \log 2 - \log(1 + e^{-w}), \quad \forall w \in \mathbb{R}_+$$

pour que $\delta_\psi \in C$ domine δ_0 . On note qu’on a bien que $\lim_{w \rightarrow \infty} \log 2 - \log(1 + e^{-w}) = \log 2$. Les risques de δ_U et δ_0 sont illustrés à la figure 1.2. On remarque que le gain maximal se produit lorsque $\theta \approx 0,645$ pour la règle de Bayes généralisée. Il est toutefois moins impressionnant que celui de l’Exemple 1.4.7 avec une amélioration d’environ 9%.

Comme au Théorème 1.4.8, on peut obtenir une condition similaire dans le cas de la perte quadratique puisque $\rho'(t) = 2t$.

1.4. MÉTHODE DE KUBOKAWA

Théorème 1.4.10 (Théorème 2 a) de [56]). *Pour la perte quadratique et $\theta \geq \theta_0$ avec θ_0 connu, l'estimateur $\delta_\psi(X) = X - \psi(X)$ satisfaisant à la condition (b) domine le MEE $X - c$ lorsque*

$$\mathbb{E}_0(X \mid X \leq x - \theta_0) \leq \psi(x), \quad \forall x$$

si $\mathbb{E}_0(|X|) < \infty$ pour tout $n \in \mathbb{N}$.

Démonstration. On reprend l'équation (1.31) et on trouve, pour tout x ,

$$\mathbb{E}_0(2(X - \psi_U(x)) \mid X \leq x - \theta_0) = 0 \Leftrightarrow \mathbb{E}_0(X \mid X \leq x - \theta_0) = \psi_U(x).$$

On isole ψ dans l'inégalité (1.28) et on déduit que

$$\psi(x) \geq \mathbb{E}_0[X \mid X \leq x - \theta_0] = \psi_U(x), \quad \forall x$$

est une condition suffisante pour que δ_ψ domine X . □

1.4.2 Pour les lois discrètes

Dans cette sous-section, nous illustrons la méthode de Kubokawa pour deux lois discrètes importantes, les lois de Poisson et binomiale négative. Pour chacun de ces modèles, on désire estimer la moyenne θ en présence d'une contrainte de la forme $\theta \geq \theta_0 > 0$. Notamment, on veut trouver des améliorations sur un estimateur cible sous la perte quadratique pondérée $\omega(\theta)(\delta - \theta)^2$ avec $\omega > 0$. Comme pour les lois continues, on recherche une classe d'estimateurs $C = \{\delta_\psi \mid \delta_\psi(X) = \delta_0(X) - \psi(X)\}$ où $\delta_0(X)$ est minimax. Puisqu'il n'y a pas de notion d'invariance aux changements de position pour les lois considérées, il sera nécessaire de cibler un estimateur minimax comme point de repère. On observera que, dans le cas de l'estimation de la moyenne d'une binomiale négative, l'estimateur cible n'est pas un ESB de θ .

Les développements de cette sous-section sont similaires à ceux de la précédente et sont inspirés de ceux présentés par Strawderman et van Eeden (pp. 58-59 de [51]). Pour les v.a. discrètes, les différences finies auront un rôle similaire à celui de la dérivée qu'on retrouve dans les développements pour les loi continues.

1.4. MÉTHODE DE KUBOKAWA

Lemme 1.4.11. *Soit une fonction ψ telle que $\lim_{w \rightarrow \infty} \psi(w) = 0$. Pour deux estimateurs de θ , $\delta_0(X)$ et $\delta_\psi(X) = \delta_0(X) - \psi(X)$, et la fonction de perte $\rho(\delta - \theta) = (\delta - \theta)^2$, on a que*

$$(\delta_\psi(X) - \theta)^2 - (\delta_0(X) - \theta)^2 = -2 \sum_{i \geq X} (\psi(i+1) - \psi(i)) \left(\delta_0(X) - \theta - \frac{\psi(i) + \psi(i+1)}{2} \right).$$

Démonstration. On réécrit, en utilisant i comme indice de la somme,

$$\begin{aligned} (\delta_\psi(X) - \theta)^2 - (\delta_0(X) - \theta)^2 &= \sum_{i \geq X} (\delta_0(X) - \theta - \psi(i))^2 - (\delta_0(X) - \theta - \psi(i+1))^2 \\ &= -2 \sum_{i \geq X} (\psi(i+1) - \psi(i)) \left(\delta_0(X) - \theta - \frac{\psi(i) + \psi(i+1)}{2} \right) \end{aligned}$$

et on obtient le résultat recherché. \square

On a noté à l'Exemple 1.3.14 avec la loi de Poisson que, bien que minimax, l'ESB est un estimateur inadmissible de la moyenne lorsque celle-ci est bornée inférieurement. Mandelkern [29] décrit un contexte similaire en physique : la recherche de signal lorsqu'on ne peut seulement mesurer le mélange de signal et de bruit de fond. Comme il est généralement possible de mesurer ce dernier isolé, on trouve une borne inférieure pour la moyenne. De plus, on trouve que l'estimateur minimax d'une binomiale négative est lui aussi inadmissible en présence d'une borne inférieure de l'espace paramétrique. Ainsi, pour $X \in \mathbb{N}$ de fonction de masse f_θ , on cherche un estimateur $\delta_\psi(X) \in \mathcal{C}$ dominant un estimateur cible $\delta_0(X)$ sous la perte quadratique.

Théorème 1.4.12. *Pour la perte quadratique et X , à valeurs sur \mathbb{N} , de fonction de masse f_θ , on a que $\delta_\psi(X) = \delta_0(X) - \psi(X)$ avec ψ une fonction non décroissante sur \mathbb{N} telle que $\lim_{w \rightarrow \infty} \psi(w) = 0$ domine l'estimateur minimax $\delta_0(X)$ lorsque*

$$\sup_{\theta \geq \theta_0} \mathbb{E}_\theta [\delta_0(X) - \theta \mid X \leq t] \leq \frac{\psi(t) + \psi(t+1)}{2}, \quad \forall t \in \mathbb{N}. \quad (1.33)$$

1.4. MÉTHODE DE KUBOKAWA

Démonstration. En vertu du Lemme 1.4.11, on trouve que

$$\begin{aligned} R(\delta_0, \theta) - R(\delta_\psi, \theta) &= -2 \sum_{x \geq 0} f_\theta(x) \sum_{t \geq x} (\psi(t+1) - \psi(t)) \left(\delta_0(x) - \theta - \frac{\psi(t) + \psi(t+1)}{2} \right) \\ &= -2 \sum_{t \geq 0} (\psi(t+1) - \psi(t)) \sum_{x=0}^t \left(\delta_0(x) - \theta - \frac{\psi(t) + \psi(t+1)}{2} \right) f_\theta(x). \end{aligned}$$

On a que l'estimateur $\delta_\psi(X) = \delta_0(X) - \psi(X)$ domine $\delta_0(X)$ si et seulement si la différence des risques est positive, soit lorsque

$$\sum_{t \geq 0} (\psi(t+1) - \psi(t)) \sum_{x=0}^t \left(\delta_0(x) - \theta - \frac{\psi(t) + \psi(t+1)}{2} \right) f_\theta(x) \leq 0. \quad (1.34)$$

Puisque $\psi(t)$ est une fonction non décroissante, une condition suffisante pour que l'inégalité (1.34) soit vérifiée est que, pour tous $\theta \geq \theta_0$ et $t \in \mathbb{N}$,

$$\begin{aligned} &\sum_{x=0}^t \left((\delta_0(x) - \theta) - \frac{\psi(t) + \psi(t+1)}{2} \right) f_\theta(x) \leq 0, \quad \forall t \in \mathbb{N}, \theta \geq \theta_0 \\ \Leftrightarrow &\mathbb{E}_\theta \left[\delta_0(X) - \frac{\psi(t) + \psi(t+1)}{2} - \theta \mid X \leq t \right] \leq 0, \quad \forall t \in \mathbb{N}, \theta \geq \theta_0 \\ \Leftrightarrow &\sup_{\theta \geq \theta_0} \mathbb{E}_\theta [\delta_0(X) - \theta \mid X \leq t] \leq \frac{\psi(t) + \psi(t+1)}{2}, \quad \forall t \in \mathbb{N}. \quad \square \end{aligned}$$

Bien que plus restrictive, une autre condition suffisante pour que δ_ψ domine $\delta_0(X)$ est que, pour $t \in \mathbb{N}$, $\sup_{\theta \geq \theta_0} \mathbb{E}_\theta [\delta_0(X) - \theta \mid X \leq t] \leq \psi(t)$. En effet, comme ψ est une fonction non décroissante, on a que $\psi(t) \leq \frac{\psi(t) + \psi(t+1)}{2}$ pour tout $t \in \mathbb{N}$. Posons

$$\psi^*(t) = \sup_{\theta \geq \theta_0} \mathbb{E}_\theta [\delta_0(X) - \theta \mid X \leq t]. \quad (1.35)$$

De plus, sous certaines conditions, on peut assurer que l'estimateur $\delta_{\psi^*}(X) = \delta_0(X) - \psi^*(X)$ ne prend pas de valeur à l'extérieur de l'espace paramétrique.

Corollaire 1.4.13. *Lorsque $\psi^*(t) = \mathbb{E}_{\theta_0} [\delta_0(X) - \theta_0 \mid X \leq t]$, alors $\delta_{\psi^*}(t) \geq \theta_0$ pour δ_0 croissant en X et $t \in \mathbb{R}$.*

1.4. MÉTHODE DE KUBOKAWA

Démonstration. On trouve, pour $t \in \mathbb{R}$, que

$$\begin{aligned}\delta_{\psi^*}(t) &= \delta_0(t) - \mathbb{E}_{\theta_0} [\delta_0(X) - \theta_0 \mid X \leq t] \geq \theta_0 \\ &\Leftrightarrow \mathbb{E}_{\theta_0} [\delta_0(X) - \theta_0 \mid X \leq t] \leq \delta_0(t) - \theta_0.\end{aligned}\quad (1.36)$$

Comme l'inégalité (1.36) est vérifiée lorsque δ_0 est une fonction croissante, on trouve le résultat. \square

Dans les exemples de cette sous-section, on considérera les estimateurs $\delta_{\psi^*}(X)$ ainsi que $\delta_+(X) = \max\{\theta_0, \delta_0(X)\}$. De plus, lorsque δ_0 est l'EVM du problème sans contrainte, δ_+ est typiquement l'EVM pour $\theta \geq \theta_0 > 0$.

Exemple 1.4.14. *Considérons un échantillon X_0, \dots, X_n indépendants tels que $X_i \sim \mathcal{P}(\theta_i)$ avec θ_0 connu. Alors, $Y \stackrel{d}{=} \sum_{i=0}^n X_i \sim \mathcal{P}(\theta)$ où $\theta = \sum_{i=0}^n \theta_i \geq \theta_0$. De plus, on a que Y est sans biais pour θ en plus d'être minimax pour la perte quadratique normalisée (voir l'Exemple 1.3.16) et d'être l'EVM du problème d'estimation sans contrainte.*

On cherche donc des améliorations pour Y avec l'estimateur $\delta_\psi \in C$ et on compare ses performances avec celles de l'EVM $\delta_+(Y) = \max\{\theta_0, Y\}$.

On montre maintenant que l'espérance tronquée de $Y - \theta$ est décroissante en θ . Premièrement, on a que $\mathbb{E}_\theta(Y - \theta \mid Y = 0) = -\theta$. Pour un $t \in \mathbb{N}^$ fixe, on a que*

$$\mathbb{E}_\theta(Y - \theta \mid Y \leq t) = \theta \left(\frac{\sum_{w=0}^{t-1} \frac{\theta^w}{w!}}{\sum_{w=0}^t \frac{\theta^w}{w!}} - 1 \right) = \theta \left(-\frac{\frac{\theta^t}{t!}}{\sum_{w=0}^t \frac{\theta^w}{w!}} \right) = -\frac{1}{\sum_{w=0}^t \frac{\theta^{w-(t+1)}}{w!/t!}}.$$

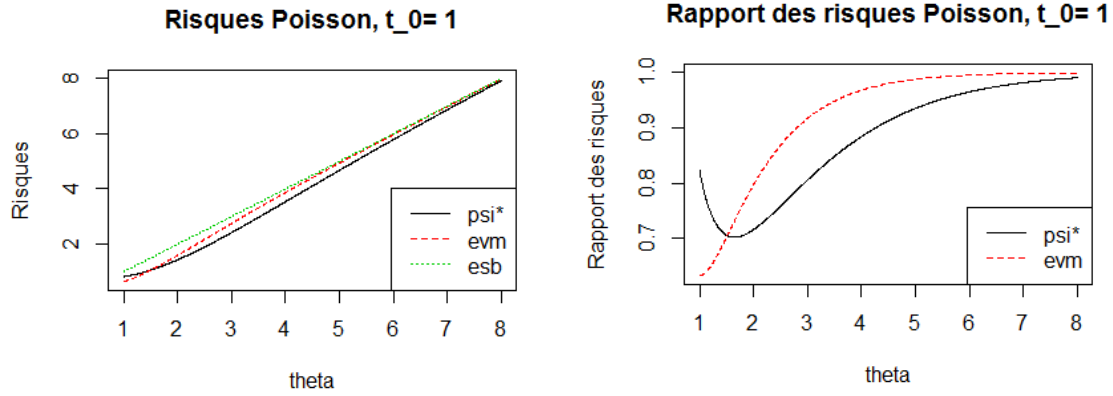
De plus, $\frac{1}{\sum_{w=0}^t \frac{\theta^{w-(t+1)}}{w!/t!}}$ est croissant en θ si et seulement si $\sum_{w=0}^t \frac{1}{\theta^{t-w}} \frac{1}{w!}$ est décroissant en θ . Comme $w \in \{0, 1, \dots, t\}$, on a que $\mathbb{E}_\theta(Y - \theta \mid Y \leq t)$ est décroissant en θ .

En appliquant le Théorème 1.4.12, on trouve que δ_ψ domine Y lorsque

$$\begin{aligned}\sup_{\theta \geq \theta_0} \mathbb{E}_\theta [\delta_0(Y) - \theta \mid Y \leq t] &= \mathbb{E}_{\theta_0} [Y - \theta_0 \mid Y \leq t] \\ &= -\frac{\theta_0^{t+1}/t!}{\sum_{n=0}^t \theta_0^n/n!} \leq \frac{\psi(t) + \psi(t+1)}{2}, \quad \forall t \in \mathbb{N}.\end{aligned}$$

Pour ψ^ tel que défini à l'équation (1.35), on trouve qu'une borne inférieure de $\delta_{\psi^*}(X)$*

1.4. MÉTHODE DE KUBOKAWA



(a) Risques de δ_{ψ^*} , δ_{EVM} et δ_0 .

(b) Rapports des risques $\frac{R(\delta_{\psi^*}, \theta)}{R(\delta_0, \theta)}$.

Figure 1.3 – Risques et rapports des risques pour la perte quadratique dans le cas d’une loi de Poisson de moyenne $\theta \geq 1$.

est θ_0 en vertu du Corollaire 1.4.13. En effet, on trouve analytiquement que

$$\delta_{\psi^*}(X) = X + \frac{\theta_0^{X+1}/X!}{\sum_{n=0}^X \theta_0^n/n!} = \begin{cases} \theta_0 & \text{si } X = 0 \\ \theta_0 + \frac{\sum_{n=0}^{X-1} (X-n)\theta_0^n/n!}{\sum_{n=0}^X \theta_0^n/n!} & \text{si } X \in \mathbb{N}^*. \end{cases}$$

Les graphiques des risques de δ_{ψ^*} , δ_{EVM} et δ_0 ainsi que de leur rapport sont à la figure 1.3 pour $\theta_0 = 1$. On remarque que le gain maximal se produit lorsque $\theta \approx 1,647$ avec une amélioration d’environ 30% et que $\delta_+(Y)$ performe mieux pour $\theta \in [1; 1,512]$.

Puisque X est une v.a. discrète sur \mathbb{N} , on ne peut utiliser l’astuce des familles de position pour simplifier la borne inférieure. Il sera donc nécessaire de démontrer que le *supremum* de l’espérance tronquée est atteint à la borne inférieure de l’espace paramétrique pour chaque distribution afin d’obtenir des résultats similaires à ceux de la sous-section 1.4.1. De plus, il n’est pas non plus possible d’associer la borne inférieure (1.35) à la règle de Bayes associée à la loi *a priori* $\pi(\theta) = \mathbb{1}_{[\theta_0, \infty)}(\theta)$. Dans les faits, il n’est même pas toujours vrai que cette dernière domine $\delta_0(X)$ pour des v.a. discrètes, comme dans le cas de la loi de Poisson qui est illustré ci-dessous.

1.4. MÉTHODE DE KUBOKAWA

Exemple 1.4.15 (p.59 de [51]). On s'intéresse au modèle $X \mid \theta \sim \mathcal{P}(\theta)$. Tel que présenté à l'Exemple 1.3.12.2, pour la loi a priori impropre $\pi(\theta) = \mathbb{1}_{[\theta_0, \infty)}(\theta)$ avec $\theta_0 > 0$, on a que la règle de Bayes généralisée est

$$\delta_U(X) = X + \frac{\theta_0^x e^{-\theta_0}}{\int_{\theta_0}^{\infty} \theta^{x-1} e^{-\theta} d\theta} = X - \phi(X)$$

où $\phi(x) = -\frac{\theta_0^x e^{-\theta_0}}{\int_{\theta_0}^{\infty} \theta^{x-1} e^{-\theta} d\theta} = -\frac{1}{\int_1^{\infty} t^{x-1} e^{-\theta_0(t-1)} dt}$. De plus, on a que ϕ est une fonction non décroissante en x et que $\lim_{x \rightarrow \infty} \phi(x) = 0$. En vertu de l'équation (1.34), on a qu'une condition équivalente pour que la règle de Bayes domine X est que

$$\sum_{t \geq 0} (\phi(t+1) - \phi(t)) \sum_{i=0}^t \left(i - \theta - \frac{\phi(t) + \phi(t+1)}{2} \right) \frac{\theta^i e^{-\theta}}{i!} \leq 0. \quad (1.37)$$

On peut montrer que, pour tout $i \in \mathbb{N}$, $\int_m^{\infty} \frac{z^i e^{-z}}{i!} dz = \sum_{j=0}^i \frac{m^j e^{-m}}{j!}$. On a donc que

$$-\frac{\phi(t) + \phi(t+1)}{2} > -\phi(t) = \frac{\theta_0^t e^{-\theta_0}}{\int_{\theta_0}^{\infty} \theta^{t-1} e^{-\theta} d\theta} = \frac{\theta_0^t e^{-\theta_0} / (t-1)!}{\sum_{j=0}^{t-1} \frac{\theta_0^j e^{-\theta_0}}{j!}}.$$

Pour $\theta = \theta_0$, on trouve, puisque $\sum_{j=0}^t \frac{\theta_0^j e^{-\theta_0}}{j!} (j - \theta_0) = -\frac{\theta_0^{t+1} e^{-\theta_0}}{t!}$, que

$$\begin{aligned} \sum_{i=0}^{t-1} \left(i - \theta_0 - \frac{\phi(t) + \phi(t+1)}{2} \right) \frac{\theta_0^i e^{-\theta_0}}{i!} &> \sum_{i=0}^{t-1} \left(i - \theta_0 + \frac{\theta_0^t e^{-\theta_0} / (t-1)!}{\sum_{j=0}^{t-1} \frac{\theta_0^j e^{-\theta_0}}{j!}} \right) \frac{\theta_0^i e^{-\theta_0}}{i!} \\ &= \frac{\theta_0^t e^{-\theta_0}}{(t-1)!} + \sum_{i=0}^{t-1} \frac{\theta_0^i e^{-\theta_0}}{i!} (i - \theta_0) \\ &= 0. \end{aligned}$$

Comme $\phi(t+1) - \phi(t) > 0$, la condition (1.37) n'est pas respectée et $R(X, \theta_0) < R(\delta_U, \theta_0)$.

Puisqu'on a démontré que X est minimax avec risque constant pour la perte $\frac{(\delta - \theta)^2}{\theta}$ lorsque $\theta \geq \theta_0 > 0$ à l'Exemple 1.3.16, la règle de Bayes généralisée δ_U ne peut l'être. Trouver une règle de Bayes admissible et minimax lorsque $X \mid \theta \sim \mathcal{P}(\theta)$ avec $\theta \geq \theta_0 > 0$ reste à ce jour un problème ouvert dans la littérature [51].

1.4. MÉTHODE DE KUBOKAWA

De plus, bien que nous considérons la perte quadratique, un estimateur sans biais de la moyenne θ , n'est pas toujours la meilleure cible à considérer. En effet, il n'est pas toujours vrai que ce dernier est minimax, comme dans le cas de la binomiale négative.

Exemple 1.4.16 (#12 p. 96 de [8]). *Considérons un échantillon X_1, \dots, X_n indépendants tels que $X_i \sim \mathcal{BN}\left(a_i, \frac{a}{a+\theta}\right)$ où $a = \sum_{i=1}^n a_i$. Alors, on trouve que $Y \stackrel{d}{=} \sum_{i=1}^n X_i \sim \mathcal{BN}\left(a, \frac{a}{a+\theta}\right)$. Ainsi, on a que sa moyenne est $\mathbb{E}_\theta(Y) = \theta$ et sa variance, $\text{Var}_\theta(Y) = \frac{\theta(a+\theta)}{a}$.*

On considère ici la perte quadratique pondérée $\rho_\omega(d - \theta) = \frac{(d-\theta)^2}{\theta(\theta+a)}$. Lorsque $\pi(\theta) = \mathbb{1}_{(0,\infty)}(\theta)$, on obtient que $\pi(\theta \mid Y = x) \propto \left(\frac{\theta}{a+\theta}\right)^x \left(\frac{a}{\theta+a}\right)^a \mathbb{1}_{(0,\infty)}(\theta)$ et que la règle de Bayes généralisée est $\delta_U(Y) = Y$ en vertu de la Proposition 1.3.6. On remarque que $\theta/a \mid Y = x \sim \mathcal{B}_\Pi(x+1, a+1)$. De plus, Y étant sans biais pour θ , on trouve que $R(\delta_U, \theta) = \frac{\text{Var}_\theta(Y)}{\theta(a+\theta)} = \frac{1}{a}$.

De même, lorsqu'on considère la loi a priori $\pi(\theta) = \left(\frac{\theta}{a+\theta}\right)^{\alpha-1} \left(\frac{a}{a+\theta}\right)^{\beta+1} \mathbb{1}_{(0,\infty)}(\theta)$ où $\alpha > 1$, la règle de Bayes associée est $\delta_{\alpha,\beta}(Y) = \frac{a}{a+\beta+1}(Y + \alpha - 1)$. Le risque associé à cette règle est $R(\delta_{\alpha,\beta}, \theta) = \frac{a}{(a+\beta+1)^2} + \frac{1}{\theta(a+\theta)} \left(\frac{a(\alpha-1) - \theta(\beta+1)}{a+\beta+1}\right)^2$.

Comme $\lim_{\beta \rightarrow 0} \delta_{1,\beta} = \frac{a}{a+1}Y$, il est plausible que $\delta_E(Y) = \frac{a}{a+1}Y$ soit une règle de Bayes par extension. On trouve que le risque associé à δ_E est

$$R(\delta_E, \theta) = \lim_{\beta \rightarrow 0} R(\delta_{1,\beta}, \theta) = \frac{a}{(a+1)^2} + \frac{\theta}{(a+\theta)(a+1)^2} < \frac{1}{a+1},$$

ce qui implique que δ_E domine δ_U et que δ_U ne peut être minimax. Pour montrer que δ_E l'est, il suffit de montrer que $r(\delta_{1,\beta}, \pi) \rightarrow \frac{1}{a+1}$ lorsque $\beta \rightarrow 0$ en vertu du Théorème 1.3.13. On a que le risque intégré de Bayes est

$$r(\delta_{1,\beta}, \pi) = \mathbb{E}^\pi \left[\frac{\text{Var}_\theta \left(\frac{aY}{a+\beta+1} \right) + \left(\frac{a\theta}{a+\beta+1} - \theta \right)^2}{\theta(\theta+a)} \right] = \frac{1}{a+\beta+1} \rightarrow \frac{1}{a+1} \text{ pour } \beta \rightarrow 0.$$

Contrairement au cas de la loi de Poisson, il n'a pas été démontré que l'estimateur minimax du problème d'estimation sans contrainte le demeure lorsque $\theta \geq \theta_0 > 0$, mais on le suppose pour la suite de cet exemple. Une communication écrite de Bill Strawderman et Éric Marchand ainsi que des évaluations numériques suggèrent que

1.4. MÉTHODE DE KUBOKAWA

le caractère minimax est vérifié.

On cherche donc des améliorations avec l'estimateur δ_ψ . On a que l'espérance tronquée de $\delta_0 - \theta$ se réécrit

$$\mathbb{E}_\theta \left[\frac{a}{a+1} Y - \theta \mid Y \leq t \right] = \frac{a}{a+1} \left[\frac{\sum_{x=0}^{t-1} \Gamma(a+1+x) (\theta/(a+\theta))^{x+1} / x!}{\sum_{x=0}^t \Gamma(a+x) (\theta/(a+\theta))^x / x!} \right] - \theta.$$

Contrairement à l'Exemple 1.4.14, il n'existe pas de forme close évidente permettant de simplifier l'expression de l'espérance tronquée pour tout $a > 0$, augmentant de ce fait le niveau de difficulté pour démontrer que cette expression est décroissante en θ .

Néanmoins, on peut démontrer que l'espérance tronquée est décroissante en θ lorsque $a = 1$, c'est-à-dire pour le cas géométrique.

Lemme 1.4.17. Soit $X \sim \text{Géo} \left(\frac{1}{1+\theta} \right)$. Alors, $\mathbb{E}_\theta(X/2 - \theta \mid X \leq t)$ est décroissant en θ .

Démonstration. On remarque premièrement que, si $\mathbb{E}_\theta(X - \theta \mid X \leq t)$ est décroissant en θ , alors $\mathbb{E}_\theta(X/2 - \theta \mid X \leq t)$ l'est aussi. Ainsi, on a que

$$\begin{aligned} \mathbb{E}_\theta(X - \theta \mid X \leq t) &= \frac{\mathbb{E}_\theta(X - \theta) - \mathbb{P}_\theta(X \geq t+1) \mathbb{E}_\theta(X \mid X \geq t+1)}{\mathbb{P}(X \leq t)} \\ &= -(t+1) \frac{\left(\frac{\theta}{1+\theta} \right)^{t+1}}{1 - \left(\frac{\theta}{1+\theta} \right)^{t+1}} \end{aligned}$$

est décroissant en θ , ce qui démontre le résultat. \square

Théorème 1.4.18. Soit $X \sim \text{Géo} \left(\frac{1}{1+\theta} \right)$ avec $\theta \geq \theta_0 > 0$. Alors, pour ψ^* tel que défini à l'équation (1.35), l'estimateur

$$\delta_{\psi^*}(X) = X/2 - \frac{\theta_0}{2} \frac{1 - \left(\frac{\theta_0}{1+\theta_0} \right)^X (1 + X/(1+\theta_0))}{1 - \left(\frac{\theta_0}{1+\theta_0} \right)^{X+1}} + \theta_0$$

domine $\delta_0(X) = X/2$ sous la perte quadratique. De plus, $\delta_{\psi^*} \geq \theta_0$ pour tout $\theta_0 > 0$.

1.4. MÉTHODE DE KUBOKAWA

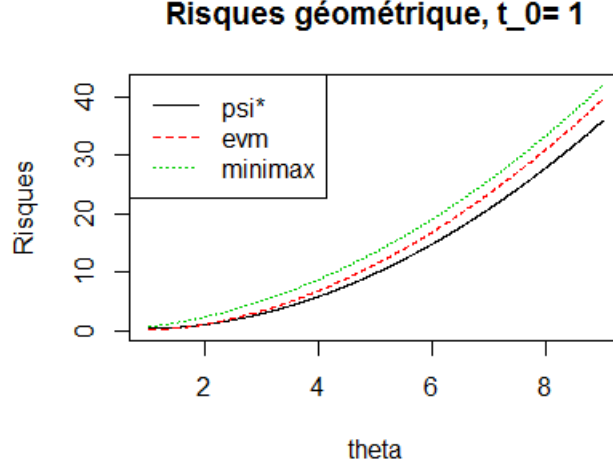


Figure 1.4 – Risques pour la perte quadratique dans le cas d’une géométrie de moyenne $\theta \geq 1$.

Démonstration. On reprend l’équation (1.35) et on trouve, en vertu du Lemme 1.4.17,

$$\begin{aligned} \psi^*(t) &= \mathbb{E}_{\theta_0}(X/2 - \theta_0 \mid X \leq t) = \frac{\theta_0 \frac{d}{d\theta_0} \left[\sum_{i=0}^X (\theta_0/(1+\theta_0))^i \right]}{2 \sum_{i=0}^X (\theta_0/(1+\theta_0))^i} - \theta_0 \\ &= \frac{\frac{X}{2} (\theta_0/(1+\theta_0))^{X+1} + (\theta_0/(1+\theta_0))^X - 1}{1 - (\theta_0/(1+\theta_0))^{X+1}} - \theta_0 \end{aligned}$$

et $\delta_{\psi^*}(X) = X/2 - \psi^*(X)$ domine $\frac{X}{2}$. Ainsi, en vertu du Corollaire 1.4.13, $\delta_{\psi^*} \geq \theta_0$. \square

Exemple 1.4.19 (Suite de l’Exemple 1.4.16). *Pour la loi géométrique, c’est-à-dire pour $a = 1$. Lorsqu’on considère $\psi^*(t) = \mathbb{E}_{\theta_0}[Y/2 - \theta_0 \mid Y \leq t]$ pour $\theta_0 = 1$, on obtient le graphique des risques de δ_{ψ^*} , δ_+ et Y à la figure 1.4.*

Chapitre 2

Estimation pour des modèles de mélange

Les modèles de mélange sont omniprésents en statistiques, on peut notamment penser aux tests d'hypothèses [26] ainsi qu'à la modélisation de données complexes [47], [16], et l'estimation de paramètres de tels modèles demeure un problème fondamental du domaine. En particulier, de tels mélanges incluent les lois décentrées bêta, chi-deux et Fisher ainsi que les distributions du R^2 ou de $\frac{R^2}{1-R^2}$ pour des données issues d'une loi normale multivariée.

Les modèles composites avec des seuils aléatoires présentent aussi leur lot de défis. Ceux-ci sont notamment utilisés en actuariat où la modélisation des réclamations d'assurances demeure un sujet de recherche important. Pigeon et Denuit [40] présentent un modèle composite lognormale-Pareto avec un seuil issu d'une gamma ou d'une lognormale afin de simuler les réclamations d'assurances non vie qui peuvent être relativement petites ou extrêmement grandes.

Dans plusieurs de ces problèmes, les estimateurs sans biais (ESB) prennent des valeurs à l'extérieur de l'espace paramétrique, les rendant de ce fait inadmissibles. Par exemple, pour $W \sim \chi_p^2(\theta)$, $W - p$ est sans biais pour θ mais représente un estimateur non plausible. Dans ce contexte, Gordon et Hall [11] proposent une solution non paramétrique basée sur le *bootstrap*. Une autre stratégie est l'amélioration d'estimateurs telle que présentée par Kubokawa, Robert et Saleh [20] ainsi que Kubokawa, Mar-

2.1. POUR UN ESPACE PARAMÉTRIQUE NON BORNÉ

chard et Strawderman [23]. Celle-ci permet notamment de déduire des estimateurs lisses et prenant des valeurs uniquement dans l'espace paramétrique.

Les résultats de ce chapitre reprennent ceux de [23] et quelques résultats originaux, tel que le Théorème 2.1.6, sont présentés. Un élément clé pour leur obtention est le rapport de vraisemblance monotone (RVM), présenté à la sous-section 1.2.2, qui permet de simplifier les expressions de risques et d'obtenir des résultats de dominance. De plus, on a recours à la méthode de Kubokawa présentée à la section 1.4 afin d'exprimer la différence des fonctions de perte comme une intégrale. Les principaux exemples considérés dans ce chapitre sont des mélanges de gamma ou de bêta de type II avec les lois de Poisson ou gamma comme paramètres de mélange en plus d'être accompagnés d'un graphique des fonctions de risque. De plus, un développement analogue pour le problème d'estimation avec une borne inférieure de l'espace paramétrique est présenté.

Pour une densité de mélange sur \mathbb{R}_+ , on propose dans ce chapitre des améliorations de l'ESB avec une approche unique pour les mélanges dénombrables et indénombrables afin d'estimer un paramètre $\theta \in \mathbb{R}_+$. Les développements s'appliquent aussi pour une densité de mélange sur \mathbb{R} si on effectue quelques ajustements, principalement au niveau des bornes des intégrales. Ce chapitre se sépare en deux sections : on considère premièrement le cas où l'espace paramétrique n'a pas de contrainte puis lorsque ce dernier est borné inférieurement.

2.1 Pour un espace paramétrique non borné

On considère un résumé stochastique $W \in \mathbb{R}_+$ issu d'une distribution de mélange admettant, pour $\theta \in [0, \infty)$, la représentation suivante

$$\begin{aligned} W \mid K = k &\sim f_k \\ K &\sim p_\theta, \end{aligned} \tag{2.1}$$

où K est de support $\mathbb{K} \subseteq [0, \infty)$, $\{f_k \mid k \in \mathbb{K}\}$ est une famille de densités sur \mathbb{R}_+ avec un RVM croissant en W avec paramètre k et la densité marginale $f(w; \theta)$ est positive pour tous w, θ . On cherche des améliorations de l'ESB δ_0 , prenant des valeurs

2.1. POUR UN ESPACE PARAMÉTRIQUE NON BORNÉ

à l'extérieur de l'espace paramétrique, à l'aide de l'estimateur

$$\delta_\psi(W) = \delta_0(W) - \psi(W) \quad (2.2)$$

où la correction ψ satisfait aux conditions suivantes :

- (a) non décroissante et absolument continue ;
- (b) $\lim_{w \rightarrow \infty} \psi(w) = 0$.

Avant de développer la différence des fonctions de risques des estimateurs δ_0 et δ_ψ , on présente quelques résultats préliminaires. On considère premièrement l'espérance conditionnelle de l'ESB

$$h(k) = \mathbb{E}(\delta_0(W) \mid K = k) = \int_0^\infty \delta_0(w) f_k(w) dw. \quad (2.3)$$

Cette fonction sera utilisée à plusieurs reprises dans les démonstrations de cette section afin de tirer avantage de la représentation de mélange. Dans ce chapitre, on suppose que $\delta_0(W)$ est une fonction strictement croissante en W , ainsi l'espérance conditionnelle $h(k)$ l'est en k en vertu des lemmes 1.2.2 et 1.2.5. De plus, on pose $\inf\{k \in \mathbb{K}\} = 0$ et on suppose que la densité p_θ est dégénérée lorsque $\theta \rightarrow 0$.

Lemme 2.1.1. *Soit W issu du modèle (2.1). Lorsque K est une v.a. discrète, on a que $h(0) = 0$.*

Démonstration. Comme δ_0 est un ESB de θ , on a que $\theta = \mathbb{E}_\theta^K(h(K)) = \sum_{k \geq 0} p_\theta(k) h(k)$. Puisque K est une v.a. dégénérée en 0 lorsque $\theta = 0$ et on trouve $0 = \sum_{k \geq 0} h(k) p_0(k) = h(0)$. \square

Ainsi, $h(k)$ est une fonction non négative lorsque K est discret. Dans le cas où K est continu, on peut obtenir un résultat similaire lorsqu'on pose des conditions sur le mélange étudié.

Lemme 2.1.2. *Soient W issu du modèle (2.1) et une fonction ξ telle que $\mathbb{E}(\xi(W) \mid K = k) = ak + b$ et que $\mathbb{E}_\theta^K(K) = c\theta$ pour $a, c > 0$. Alors $\delta_0(W) = \frac{\xi(W) - b}{ac}$ est un ESB de θ et $\lim_{k \rightarrow 0} h(k) = 0$.*

Démonstration. Par hypothèse, on a que $\mathbb{E}_\theta(\xi(W)) = a(c\theta) + b$. On a donc qu'un ESB de θ est $\frac{\xi(W) - b}{ac}$. D'où, $h(k) = \mathbb{E}\left(\frac{\xi(W) - b}{ac} \mid K = k\right) = \frac{k}{c}$ et on obtient le résultat. \square

2.1. POUR UN ESPACE PARAMÉTRIQUE NON BORNÉ

On illustre le dernier lemme avec le mélange d'une log-normale et d'une gamma.

Exemple 2.1.3. Supposons que $W \mid K = k \sim \mathcal{LN}(\mu + k, \sigma^2)$ avec $K \sim \mathcal{G}(\theta, \beta)$ pour $\beta, \sigma^2 > 0$ et $\mu \geq 0$. On a que $\frac{\ln W - \mu}{\beta}$ est un ESB de θ puisque

$$\mathbb{E}_\theta(\ln W) = \mathbb{E}_\theta^K \mathbb{E}(\ln X \mid K = k) = \mathbb{E}_\theta(\mu + K) = \mu + \beta\theta.$$

Puisque δ_0 est une fonction strictement croissante en W , on a que $h(0) \geq 0$ en vertu du Lemme 2.1.2. En effet, on trouve analytiquement que

$$h(k) = \frac{\mathbb{E}(\ln X \mid K = k) - \mu}{\beta} = \frac{k}{\beta}.$$

Bien qu'assez simples, les lemmes ci-dessus aideront à simplifier les expressions de risques. Une autre astuce utilisée dans les développements de ce chapitre est de considérer l'écart entre l'ESB et l'espérance conditionnelle $h(k)$. On pose

$$g_k(w) = \int_0^w (h(k) - \delta_0(t)) f_k(t) dt = h(k) F_k(w) - \int_0^w \delta_0(t) f_k(t) dt. \quad (2.4)$$

Par définition, on a que $\frac{d}{dw} g_k(w) = (h(k) - \delta_0(w)) f_k(w)$, une propriété qui sera mise à profit notamment en ayant recours à la méthode de Kubokawa.

Remarque 2.1.4. Lorsque $\delta_0(W) = c_0 W - d_0$ avec $c_0 > 0$ et $d_0 \geq 0$, $\frac{d}{dw} g_k(w)$ aura exactement un changement de signe en $\frac{h(k) - d_0}{c_0}$.

Lemme 2.1.5. La fonction g_k telle que définie à l'équation (2.4) a les propriétés suivantes :

- i) $\lim_{w \rightarrow 0} g_k(w) = \lim_{w \rightarrow \infty} g_k(w) = 0$ pour tout $k \in \mathbb{K}$;
- ii) $g_k(w) > 0$ pour tous $w \in \mathbb{R}_+$ et $k \in \mathbb{K}$;

Démonstration. Soit $k \in \mathbb{K}$.

On montre premièrement i). Puisque $F_k(0) = 0$ pour tout k , alors $\lim_{w \rightarrow 0} g_k(w) = h(k) F_k(0) = 0$. De plus, on trouve de l'équation (2.4) que

$$\lim_{w \rightarrow \infty} g_k(w) = h(k) - \int_0^\infty \delta_0(t) f_k(t) dt = h(k) - h(k) = 0.$$

2.1. POUR UN ESPACE PARAMÉTRIQUE NON BORNÉ

Ensuite, pour [ii](#)), on a que

$$\begin{aligned} 0 < g_k(w) &\Leftrightarrow 0 < \int_0^\infty \delta_0(t) f_k(t) dt - \frac{\int_0^w \delta_0(t) f_k(t) dt}{F_k(w)} \\ &= \mathbb{E}(\delta_0(W) \mid K = k) - \mathbb{E}(\delta_0(W) \mid K = k, W \leq w). \end{aligned}$$

et, comme on suppose que δ_0 est une fonction strictement croissante, l'espérance conditionnelle $\mathbb{E}(\delta_0(W) \mid K = k, W \leq w)$ l'est aussi en w en vertu des lemmes [1.2.2](#) et [1.2.6](#). \square

On peut maintenant énoncer les résultats principaux de cette section. Ces derniers généralisent ceux de Kubokawa et coll. [\[23\]](#) qui ont démontré l'énoncé suivant lorsque $\delta_0(W) = c_0 W - d_0$ est un ESB de θ pour des constantes $c_0 > 0$ et $d_0 \geq 0$ pour W un mélange dénombrable. En relâchant les hypothèses sur l'ESB et sur la représentation de mélange de W , on ne peut autant simplifier l'expression finale de la borne inférieure de ψ .

Théorème 2.1.6. *Soient $W \in \mathbb{R}_+$ issu du modèle [\(2.1\)](#) et $\delta_0(W)$ un ESB de θ strictement croissant en W tels que*

$$\frac{g_{k+1}(w)}{g_k(w)} \geq \frac{f_{k+1}(w)}{f_k(w)}, \quad \forall w > 0, k \in \mathbb{K}. \quad (2.5)$$

Alors, pour estimer $\theta \in [0, \infty)$, δ_0 est dominé par $\delta_\psi(W) = \delta_0(W) - \psi(W)$ où ψ est absolument continue, non décroissante, avec $\lim_{w \rightarrow \infty} \psi(w) = 0$ sous la perte quadratique lorsque

$$\psi(w) \geq \mathbb{E}(\delta_0(W) \mid W \leq w, K = 0) - h(0) = \psi^*(w), \quad \forall w > 0 \quad (2.6)$$

si $\mathbb{E}_\theta(\delta_0^2(W)), \mathbb{E}_\theta(\psi^2(W)) < \infty$, $f(w; \theta) > 0$ pour tous w, θ et $f_0 \neq 0$.

Remarque 2.1.7. *Lorsque $f_k(w) = c_k(u(w))^k f_0(w)$ avec u non décroissante puis f_0 , u différentiables, Kubokawa et coll. [\[23\]](#) ont démontré à la Remarque 2.1 que la condition [\(2.5\)](#) est vérifiée. Cette représentation regroupe notamment les distributions de la famille exponentielle.*

2.1. POUR UN ESPACE PARAMÉTRIQUE NON BORNÉ

Remarque 2.1.8. Lorsque $h(0) \geq 0$, on trouve que l'estimateur $\delta_{\psi^*}(W) = \delta_0(W) - \psi^*(W)$ est non négatif. En effet, on a que

$$\delta_{\psi^*}(w) = \delta_0(w) - \mathbb{E}(\delta_0(W) \mid W \leq w, K = 0) + h(0) \geq h(0).$$

Démonstration du Théorème 2.1.6. On se base sur la démonstration du Théorème 2.1 de Kubokawa et coll. [23]. On a que la différence des risques entre δ_0 et δ_ψ est

$$R(\delta_0, \theta) - R(\delta_\psi, \theta) = 2\mathbb{E}_\theta[\psi(W)(\delta_0(W) - \theta)] - \mathbb{E}_\theta[\psi^2(W)]. \quad (2.7)$$

On développe le premier terme et on trouve, pour $\mathbb{E}_\theta^K(\cdot)$ l'espérance sur K ,

$$\begin{aligned} \mathbb{E}_\theta[\psi(W)(\delta_0(W) - \theta)] &= \mathbb{E}_\theta^K \mathbb{E}[\psi(W)(\delta_0(W) - h(k)) \mid K = k] \\ &\quad + \mathbb{E}_\theta^K \mathbb{E}[\psi(W)(h(k) - \theta) \mid K = k]. \end{aligned}$$

On intègre par partie la première espérance conditionnelle et on trouve

$$\begin{aligned} \mathbb{E}_\theta^K \mathbb{E}[\psi(W)(\delta_0(W) - h(k)) \mid K = k] &= \mathbb{E}_\theta^K \left[\int_0^\infty \psi(t) \frac{d}{dt} (-g_k(t)) dt \right] \\ &= \mathbb{E}_\theta^K \left[\int_0^\infty \psi'(t) g_k(t) dt \right] \\ &= \int_0^\infty \psi'(t) g(t; \theta) dt, \end{aligned} \quad (2.8)$$

où $g(t; \theta) = \mathbb{E}_\theta^K(g_K(t))$. En vertu de l'inégalité des covariances (voir Lemme 6.6 (b) de [25]), on a aussi que

$$\mathbb{E}_\theta^K \mathbb{E}[\psi(W)(h(k) - \theta) \mid K = k] \geq \mathbb{E}_\theta[\psi(W)] \mathbb{E}_\theta^K[(h(K) - \theta)] = 0 \quad (2.9)$$

puisque $\mathbb{E}_\theta(\psi(W) \mid K = k)$ et $h(k)$ sont croissantes en k ainsi que $\mathbb{E}_\theta^K(h(K)) = \theta$. Pour le second terme de l'équation (2.7), on applique la méthode de Kubokawa et on

2.1. POUR UN ESPACE PARAMÉTRIQUE NON BORNÉ

trouve, pour $f(w; \theta)$ la densité de W par rapport au paramètre θ ,

$$\begin{aligned} \mathbb{E}_\theta(\psi^2(W)) &= \int_0^\infty \psi^2(w) f(w; \theta) dw \\ &= \int_0^\infty \int_w^\infty -\frac{d}{dt} \psi^2(t) dt f(w; \theta) dw \\ &= -2 \int_0^\infty \psi'(t) \psi(t) \int_0^t f(w; \theta) dw dt. \end{aligned} \quad (2.10)$$

Ainsi, en appliquant les simplifications (2.8), (2.9) et (2.10) dans l'équation (2.7), on trouve que la différence des risques est

$$R(\delta_0, \theta) - R(\delta_\psi, \theta) \geq 2 \int_0^\infty \left[g(t; \theta) + \psi(t) \int_0^t f(w; \theta) dw \right] \psi'(t) dt.$$

Comme ψ est une fonction non décroissante, une condition suffisante pour que la différence des risques soit positive est que

$$\begin{aligned} g(w; \theta) + \psi(w) \int_0^w f(t; \theta) dt &\geq 0, \quad \forall w > 0, \theta \geq 0 \\ \Leftrightarrow -\inf_{\theta \geq 0} \frac{g(w; \theta)}{\int_0^w f(t; \theta) dt} &\leq \psi(w), \quad \forall w > 0. \end{aligned} \quad (2.11)$$

Par la condition ii) du Lemme 2.1.5 et l'expression (2.5), on a que, pour tous $w > 0$ et $k \in \mathbb{K}$, $g_k(w) f_0(w) \geq g_0(w) f_k(w)$. En utilisant le fait que f_k a un RVM croissant en W avec paramètre k et en appliquant l'espérance sur K de chaque côté de l'inégalité, on trouve, pour $t \in (0, w)$, que $g(w; \theta) f_0(t) \geq f(t; \theta) g_0(w)$. D'où, pour tous $\theta \geq 0$ et $w \in \mathbb{R}_+$,

$$g(w; \theta) \int_0^w f_0(t) dt \geq g_0(w) \int_0^w f(t, \theta) dt \Leftrightarrow \frac{g(w; \theta)}{\int_0^w f(t; \theta) dt} \geq \frac{g_0(w)}{\int_0^w f_0(t) dt}.$$

Ainsi, une condition suffisante pour que δ_ψ domine δ_0 est que

$$\psi(w) \geq \frac{\int_0^w \delta_0(t) f_0(t) dt}{\int_0^w f_0(t) dt} - h(0) = \mathbb{E}(\delta_0(W) \mid W \leq w, K = 0) - h(0), \quad \forall w > 0. \quad \square$$

On illustre premièrement ce résultat avec le mélange d'une loi lognormale et d'une gamma.

2.1. POUR UN ESPACE PARAMÉTRIQUE NON BORNÉ

Exemple 2.1.9 (Suite de l'Exemple 2.1.3). Pour $W \mid K = k \sim \mathcal{LN}(\mu + k, \sigma^2)$ avec $K \sim \mathcal{G}(\theta, \beta)$ où $\beta, \sigma > 0$ et $\mu \geq 0$ connus. On remarque que f_k a un RVM croissant en W avec paramètre k et que $\delta_0(W) = \frac{\ln W - \mu}{\beta}$ est un ESB de θ en plus d'être strictement croissant en W en vertu des propriétés des logarithmes. De plus, on trouve que $\mathbb{E}_\theta(\delta_0^2(W)) = \frac{\sigma^2}{\beta^2} + \theta(\theta + 1)$. On obtient les quantités suivantes :

$$h(k) = \mathbb{E} \left(\frac{\ln W - \mu}{\beta} \mid K = k \right) = \frac{k}{\beta};$$

$$g_k(t) = \int_0^t \left[\frac{k - (\ln t - \mu)}{\beta} \right] \frac{1}{t\sigma\sqrt{2\pi}} e^{-(\ln t - \mu - k)^2 / 2\sigma^2} dt = \frac{\sigma^2}{\beta} \phi \left(\frac{\ln t - \mu}{\sigma} \right);$$

où ϕ est la densité d'une $\mathcal{N}(0, 1)$ et Φ sa répartition. En appliquant le Théorème 2.1.6, on a que $\delta_\psi(W) = \delta_0(W) - \psi(W)$ satisfaisant aux conditions (a) et (b) domine l'ESB sous la perte quadratique lorsque

$$\psi(w) \geq \frac{-g_0(w)}{\int_0^w f_0(t)dt} = -\frac{\sigma^2}{\beta} \frac{\phi \left(\frac{\ln w - \mu}{\sigma} \right)}{\Phi \left(\frac{\ln w - \mu}{\sigma} \right)} = \psi^*(w), \quad \forall w > 0.$$

Si on considère $\psi(w) \equiv \psi^*(w)$, on remarque premièrement que $\lim_{w \rightarrow \infty} \psi(w) = 0$. De plus, comme la loi normale est log-concave, alors le rapport $\frac{\phi(z)}{\Phi(z)}$ est croissant en $z \in \mathbb{R}$. Ainsi, δ_{ψ^*} domine δ_0 sous la perte quadratique.

De retour au cas général, comme on suppose $\delta_0(W)$ strictement croissant en W au Théorème 2.1.6, l'espérance conditionnelle $\mathbb{E}(\delta_0(W) \mid W \leq w, K = 0)$, l'est aussi en w en vertu des lemmes 1.2.2 et 1.2.6. De l'expression (2.6), on peut déduire plusieurs fonctions ψ telles que l'estimateur δ_ψ domine δ_0 dont :

- l'enveloppe inférieure $\psi^*(w) = \mathbb{E}(\delta_0(W) \mid W \leq w, K = 0) - h(0)$;
- $\psi_+(w) = \min\{0, \delta_0(w)\}$ et on trouve que $\delta_{\psi_+}(W) = \max\{0, \delta_0(W)\}$.

Dans les exemples qui suivent, on a que $\mathbb{E}_\theta(W)$ est une fonction affine de θ . Ceci amène les simplifications suivantes.

Corollaire 2.1.10. Soit W issu du modèle (2.1) tel que $\delta_0(W) = c_0 W - d_0$ est sans biais pour θ avec $c_0 > 0$ et $d_0 \geq 0$ et que l'expression (2.5) est vérifiée. Sous la perte

2.1. POUR UN ESPACE PARAMÉTRIQUE NON BORNÉ

quadratique, δ_ψ tel que défini à l'équation (2.2) domine δ_0 lorsque

$$\psi(w) \geq \frac{\int_0^w c_0 t f_0(t) dt}{\int_0^w f_0(t) dt} - (h(0) + d_0), \quad \forall w > 0$$

où $0 = \inf\{k \in \mathbb{K}\}$ si $\mathbb{E}_\theta(W^2), \mathbb{E}_\theta[\psi^2(W)] < \infty$, $f(w; \theta) > 0$ pour tous w, θ et $f_0 \neq 0$.

On illustre ce phénomène avec le mélange d'une inverse-gamma, la distribution de la réciproque d'une loi gamma, et d'une gamma. De plus, on remarquera que la densité marginale s'écrit avec une expression assez simple lorsque le paramètre de forme de la loi conditionnelle est un entier, en tirant profit du binôme de Newton. On peut notamment lier cette loi de probabilité, appelée distribution d'Erlang dans la littérature, à celle de Poisson à l'aide des processus de Poisson (voir [18] pour une présentation plus détaillée).

Exemple 2.1.11. Soit un échantillon X_1, \dots, X_n tels que $X_i \mid K = k \sim \mathcal{G}(l_i, (m + k)^{-1})$ indépendants pour $i \in \{1, \dots, n\}$ avec $K \sim \mathcal{G}(\alpha, \theta)$ où $l_i, m, \alpha > 0$ connus et $\theta > 0$. On pose $W \stackrel{d}{=} (\sum_{i=1}^n X_i)^{-1}$, alors $W \mid K = k \sim \mathcal{IG}(l, m + k)$ où $l = \sum_{i=1}^n l_i$ et sa densité s'écrit

$$\begin{aligned} f(w; \theta) &= \int_0^\infty \frac{(m + k)^l w^{-l-1} e^{-(m+k)/w}}{\Gamma(l)} \mathbf{1}_{(0, \infty)}(w) \frac{k^{\alpha-1} e^{-k/\theta}}{\Gamma(\alpha) \theta^\alpha} dk \\ &= \frac{w^{-l-1} e^{-m/w}}{\Gamma(l) \Gamma(\alpha) \theta^\alpha} \mathbf{1}_{(0, \infty)}(w) \int_0^\infty (m + k)^l k^{\alpha-1} e^{-k(1/w + 1/\theta)} dk \end{aligned}$$

On remarque que, lorsque l est un entier non négatif, $(m + k)^l = \sum_{i=0}^l \binom{l}{i} k^i m^{l-i}$. Dans ce cas, on trouve l'expression suivante

$$f(w; \theta) = \frac{w^{-l-1} e^{-mw}}{\Gamma(l) \Gamma(\alpha) \theta^\alpha} \mathbf{1}_{(0, \infty)}(w) \sum_{i=0}^l \binom{l}{i} \frac{m^{l-i} \Gamma(\alpha + i)}{(w + 1/\theta)^{\alpha+i}}$$

ce qui simplifie les évaluations numériques. De plus, pour $l > 2$, on a que $\mathbb{E}_\theta(W) = \frac{m + \alpha\theta}{l-1}$ et que $\text{Var}_\theta(W) = \frac{m^2 + 2m\alpha\theta + \alpha\theta^2 + n^2\theta^2}{(l-1)^2(l-2)} + \frac{\alpha\theta^2}{(l-1)^2}$. Dans le contexte du Théorème

2.1. POUR UN ESPACE PARAMÉTRIQUE NON BORNÉ

[2.1.6](#), on obtient les quantités suivantes :

$$h(k) = \mathbb{E} \left(\frac{(l-1)W - m}{\alpha} \mid K = k \right) = \frac{k}{\alpha}$$

$$g_k(w) = \int_0^w \left[\frac{k}{\alpha} - \frac{(l-1)t - m}{\alpha} \right] f_k(t) dt = \frac{(m/w)^{l-1} e^{-m/w}}{\Gamma(l)}.$$

On remarque qu'on a ici que $h(0) = 0$, ce qui est cohérent avec le Lemme [2.1.2](#). En vertu du Théorème [2.1.6](#), on a qu'une condition suffisante pour que l'estimateur $\delta_\psi(W) = \frac{(l-1)W - m}{\alpha} - \psi(W)$ satisfaisant aux conditions (a) et (b) domine l'ESB sous la perte quadratique est

$$\psi(w) \geq \frac{-g_0(x)}{\int_0^w f_0(t) dt} = -\frac{(m/w)^{l-1} e^{-m/w}}{\Gamma_{m/w}(l)} = \psi^*(w), \quad \forall w > 0$$

où $\Gamma_z(\alpha) = \int_z^\infty t^{\alpha-1} e^{-t} dt$.

Lorsqu'on considère $\psi(w) \equiv \psi^*(w) = -\frac{1}{\int_1^\infty u^{l-1} e^{-\frac{m}{w}(u-1)} du}$, on remarque que ψ^* est une fonction croissante en w et que $\lim_{w \rightarrow \infty} \psi^*(w) = 0$. Lorsque $l = m = n = 3$, on obtient le graphique du rapport des risques $\frac{R(\delta_{\psi^*}, \theta)}{R(\delta_0, \theta)}$ à la figure [2.1](#). On observe que le gain maximal se produit lorsque $\theta \approx 0,56$ avec une amélioration d'environ 13%.

On s'inspire du modèle développé par Pigeon et Denuit [\[40\]](#) pour le prochain exemple. Pour simplifier les calculs, on considère un modèle composite uniforme-Pareto avec un seuil issu d'une gamma.

Exemple 2.1.12. Avec la notation du modèle [\(2.1\)](#), on pose, pour $\alpha > 2$,

$$f_k(w) = \frac{\alpha}{\alpha+1} \frac{1}{k+1} \mathbf{1}_{(0, k+1)}(w) + \frac{1}{\alpha+1} \frac{\alpha(k+1)^\alpha}{w^{\alpha+1}} \mathbf{1}_{[k+1, \infty)}(w),$$

un modèle composite d'une uniforme et d'une Pareto, avec $K \sim \mathcal{G}(1, \theta)$. On peut vérifier analytiquement que la famille de densités $\{f_k \mid k \in \mathbb{R}_+\}$ a un RVM croissant en W avec paramètre k . On obtient alors les moments $\mathbb{E}_\theta(W) = (\theta+1) \left[\frac{\alpha}{2(\alpha+1)} + \frac{\alpha}{(\alpha+1)(\alpha-1)} \right]$ et que $\mathbb{E}_\theta(W^2) = (2\theta^2 + 2\theta + 1) \left[\frac{\alpha}{3(\alpha+1)} + \frac{\alpha}{(\alpha+1)(\alpha-2)} \right]$. On cherche donc des améliorations de l'ESB $\delta_0(W) = \frac{\frac{\alpha+1}{2} W}{\frac{\alpha}{2} + \frac{1}{\alpha-1}} - 1$. Dans le contexte du Théorème [2.1.6](#), on obtient

2.1. POUR UN ESPACE PARAMÉTRIQUE NON BORNÉ

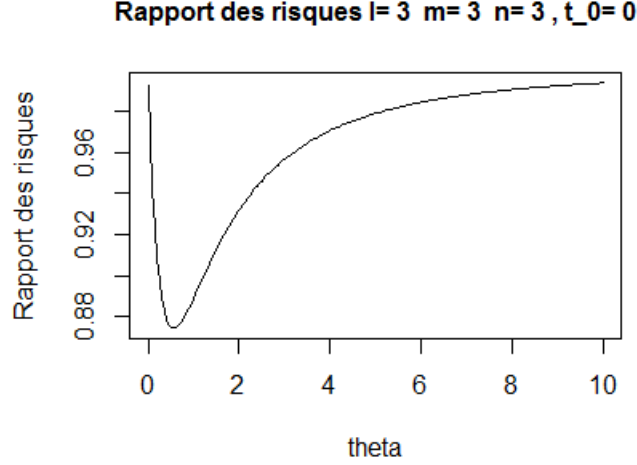


Figure 2.1 – Rapport des risques $\frac{R(\delta_{\psi^*, \theta})}{R(\delta_0, \theta)}$ pour un mélange de lois inverse-gamma et gamma d'échelle $\theta \geq 0$.

les quantités suivantes :

$$h(k) = \mathbb{E} \left(\frac{\frac{\alpha+1}{\frac{1}{2} + \frac{1}{\alpha-1}} W}{\frac{1}{2} + \frac{1}{\alpha-1}} - 1 \mid K = k \right) = k$$

$$g_k(w) = \begin{cases} w \left[\frac{\alpha}{\alpha+1} - \frac{(\alpha-1)w}{(k+1)(\alpha+1)} \right] & \text{si } 0 < w < k+1 \\ \frac{k+1}{\alpha+1} \left[\frac{\alpha-1}{\alpha+1} - \left(\frac{k+1}{w} \right)^\alpha + 2 \left(\frac{k+1}{w} \right)^{\alpha-1} \right] & \text{si } w \geq k+1. \end{cases}$$

On remarque qu'on a ici que $h(0) = 0$, ce qui est cohérent avec le Lemme 2.1.1, et on peut montrer analytiquement que l'expression (2.5) est vérifiée. En vertu du Théorème 2.1.6, on a qu'une condition suffisante pour que l'estimateur $\delta_\psi(W) = \frac{\frac{\alpha+1}{\frac{1}{2} + \frac{1}{\alpha-1}} W}{\frac{1}{2} + \frac{1}{\alpha-1}} - 1 - \psi(W)$ satisfaisant aux conditions (a) et (b) domine l'ESB sous la perte quadratique est

$$\psi(w) \geq -\frac{g_0(w)}{\int_0^w f_0(t)dt} = \psi^*(w) = \begin{cases} \frac{\alpha-1}{\alpha} w - 1 & \text{si } 0 < w < 1 \\ \frac{\frac{\alpha-1}{\alpha+1} - w^{-\alpha} + 2w^{1-\alpha}}{\alpha+1 - w^{-\alpha}} & \text{si } w \geq 1. \end{cases}$$

En général, la fonction ψ^* , la borne inférieure trouvée au Théorème 2.1.6, demeure

2.1. POUR UN ESPACE PARAMÉTRIQUE NON BORNÉ

plutôt difficile à interpréter en terme de l'espace paramétrique. On remarque notamment qu'on conditionne par rapport l'infimum de K , bien qu'on ne soit pas en mesure de faire un lien explicite avec θ sans ajouter de conditions supplémentaires.

Corollaire 2.1.13. *Soit W issu du modèle (2.1). Lorsque p_θ a un RVM croissant en K avec paramètre θ , on a que $\delta_\psi(W) = \delta_0(W) - \psi(W)$ domine l'ESB δ_0 sous la perte quadratique si*

$$\psi(w) \geq \mathbb{E}_0(\delta_0(W) \mid W \leq w) - \mathbb{E}_w^L(h(L)), \quad \forall w > 0 \quad (2.12)$$

lorsque $\mathbb{E}_\theta(\delta_0^2(W)), \mathbb{E}_\theta(\psi^2(W)) < \infty$, $f(\cdot; 0) \not\equiv 0$ et où

$$L \sim F_l(w)p_0(l)/\mathbb{E}_0^K(F_K(w)). \quad (2.13)$$

Démonstration. On reprend la démonstration du Théorème 2.1.6 jusqu'à l'inégalité (2.11). On se base ensuite sur la démonstration du Théorème 3.1 de [23].

On montre premièrement que $\frac{g(x;\theta)}{\int_0^x f(t;\theta)dt} \geq \frac{g(x;0)}{\int_0^x f(t;0)dt}$. On a que

$$\frac{g(x;\theta)}{f(x;\theta)} = \frac{\mathbb{E}_\theta^K\left(\frac{g_K(x)}{f_K(x)}f_K(x)\right)}{f(x;\theta)} = \mathbb{E}_{x,\theta}^M\left(\frac{g_M(x)}{f_M(x)}\right)$$

où $M \sim f_m(x)p_\theta(m)/f(x;\theta)$. En utilisant la même astuce qu'au Lemme 1.2.8, on a que $f_l(m)p_\theta(m)/f(x;\theta)$ a un RVM croissant en M avec paramètre θ comme p_θ en a un en K avec paramètre θ . Puisque $\frac{g_M(x)}{f_M(x)}$ est croissant en M en vertu de l'expression (2.5), on trouve que $\mathbb{E}_{x,\theta}^M\left(\frac{g_M(x)}{f_M(x)}\right)$ est croissant en $\theta \geq 0$. D'où, $\frac{g(x;\theta)}{f(x;\theta)} \geq \frac{g(x;0)}{f(x;0)}$ pour tout θ . Comme $f(\cdot; \theta)$ a un RVM croissant en W avec paramètre θ en vertu du Lemme 1.2.8, on a que

$$\frac{f(w;\theta)}{f(w;0)} = \frac{\int_0^w \frac{f(w;\theta)}{f(w;0)}f(t;0)dt}{\int_0^w f(t;0)dt} \geq \frac{\int_0^w \frac{f(t;\theta)}{f(t;0)}f(t;0)dt}{\int_0^w f(t;0)dt} = \frac{\int_0^w f(t;\theta)dt}{\int_0^w f(t;0)dt}.$$

D'où, $\frac{g(w;\theta)}{\int_0^w f(t;\theta)dt} \geq \frac{g(w;0)}{\int_0^w f(t;0)dt}$. En retravaillant sur $g(w;0)$ à l'aide de l'équation (2.4),

2.1. POUR UN ESPACE PARAMÉTRIQUE NON BORNÉ

on a que

$$g(w; 0) = \mathbb{E}_0^K(g_K(w)) = \mathbb{E}_0^K(h(K)F_K(w)) - \int_0^w \delta_0(t)f(t; 0)dt.$$

On obtient donc que

$$\psi(w) \geq \frac{-g(w; 0)}{\int_0^w f(t; 0)dt} = \frac{\int_0^w \delta_0(t)f(t; 0)dt}{\int_0^w f(t; 0)dt} - \mathbb{E}_w^L(h(L))$$

où la densité de L est donnée par (2.13). \square

Contrairement à la borne obtenue au Théorème 2.1.6, celle du dernier corollaire est plus explicite. En effet, les deux espérances considérées sont évaluées en 0, l'infimum de l'espace paramétrique. De plus, on a que $\mathbb{E}_{x,0}^M(h(M))$ est croissant en x en vertu des propriétés des RVM en plus d'être négatif comme

$$\mathbb{E}_w^L(h(L)) = \frac{\mathbb{E}_0^K(h(K)F_K(w))}{F(w; 0)} \leq \frac{\mathbb{E}_0^K(h(K))F(w; 0)}{F(w; 0)} = \mathbb{E}_0^K(h(K)) = 0.$$

D'autre part, si $h \geq 0$, alors $0 = \mathbb{E}_w^L(h(L))$ et on déduit qu'une condition suffisante afin que δ_ψ domine l'ESB est

$$\psi(w) \geq \mathbb{E}_0(\delta_0(W) \mid W \leq w), \quad \forall w > 0,$$

la borne inférieure déduite avec la méthode de Kubokawa donnée au Théorème 1.4.10.

Lorsque la loi du paramètre de mélange est dégénérée en 0 lorsque $\theta = 0$, calculer l'espérance tronquée du Corollaire 2.1.13 devient trivial. Ainsi, les bornes inférieures données au Théorème 2.1.6 et au Corollaire 2.1.13 sont exactement les mêmes.

On illustre ces résultats avec la Fisher décentrée et la fonction $\theta = \frac{\rho^2}{1-\rho^2}$, où ρ^2 est le coefficient de détermination d'une loi normale multivariée, deux mélanges de bêta de type II avec, respectivement, une loi de Poisson et une binomiale négative.

Exemple 2.1.14 (Exemple 2.2 de [23]). *Soit $W \sim F_{m,n}(\theta)$ avec $m, n > 0$ degrés de liberté et de paramètre de décentralité $\theta \geq 0$. On a que $\mathbb{E}_\theta(W) = \frac{n(m+\theta)}{m(n-2)}$ lorsque $n > 2$ et que $\text{Var}_\theta(W) = 2\frac{m^2}{n^2} \frac{(m+\theta)^2 + (m+2\theta)(n-2)}{(n-2)^2(n-4)}$ pour $n > 4$. On cherche des améliorations*

2.1. POUR UN ESPACE PARAMÉTRIQUE NON BORNÉ

de l'ESB $\frac{m(n-2)}{n}W - m$. On a que W admet la représentation suivante :

$$\begin{aligned} \frac{m}{n}W \mid K = k &\sim \mathcal{B}_{\Pi}(k + m/2, n/2) \\ K &\sim \mathcal{P}(\theta/2). \end{aligned}$$

Dans le contexte du Théorème 2.1.6, on obtient les quantités suivantes :

$$\begin{aligned} h(k) &= \mathbb{E} \left(\frac{m(n-2)}{n}W - m \mid K = k \right) = (n-2) \frac{k + (m+n)/2}{n/2 - 1} - m = 2k; \\ g_k(w) &= \int_0^w (2k - \frac{m(n-2)}{n}t + m) f_k(t) dt = 2w \left(1 + \frac{m}{n}w \right) f_k(w). \end{aligned}$$

En appliquant le Corollaire 2.1.13, on a que $\delta_\psi(W) = \frac{m(n-2)}{n}W - m - \psi(W)$ avec ψ satisfaisant aux conditions (a), (b) domine l'ESB si

$$\psi(w) \geq \frac{-g_0(w)}{\int_0^w f_0(t) dt} = -2 \frac{(\frac{m}{n}w)^{\frac{m}{2}} (1 + \frac{m}{n}w)^{-(\frac{m+n}{2}-1)}}{\mathcal{B}_{\frac{mw/n}{1+mw/n}}(\frac{m}{2}, \frac{n}{2})} = \psi^*(w), \quad \forall w > 0$$

où $\mathcal{B}_x(\alpha, \beta) = \int_0^x t^{\alpha-1} (1-t)^{\beta-1} dt$ pour $x \in [0, 1]$ est la fonction bêta incomplète.

Lorsqu'on considère $\psi(w) \equiv \psi^*(w)$, on remarque que $\lim_{w \rightarrow \infty} \psi^*(w) = 0$. Lorsque $m = 7$ et $n = 5$, on obtient le graphique du rapport des risques $\frac{\mathcal{R}(\delta_{\psi^*}, \theta)}{\mathcal{R}(\delta_0, \theta)}$ à la figure 2.2. On remarque que le gain maximal se produit lorsque $\theta \approx 4,945$ avec une amélioration d'environ 12%.

Exemple 2.1.15 (section 4 de [23]). On cherche à estimer ρ^2 le coefficient de corrélation multiple entre les variables V_1 et $\mathbf{V}_2 = (V_2, \dots, V_m)^T$ à l'aide de R^2 , son pendant échantillonnal. Une présentation plus détaillée est donnée à la sous-section 1.1.2. Pour cet exemple, on prend l'approche préconisée par Muirhead [38] où l'on estime $\frac{\rho^2}{1-\rho^2}$ à l'aide de $\frac{R^2}{1-R^2}$. Tel que présenté au Théorème 5.2.5 de [37], on a, pour une taille d'échantillon $n+1$, que $X \stackrel{d}{=} \frac{R^2}{1-R^2}$ admet la représentation suivante :

$$\begin{aligned} X \mid K = k &\sim \mathcal{B}_{\Pi} \left(\frac{m-1}{2} + k, \frac{n-m+1}{2} \right) \\ K &\sim \mathcal{BN}(n/2, (1+\theta)^{-1}). \end{aligned}$$

2.1. POUR UN ESPACE PARAMÉTRIQUE NON BORNÉ

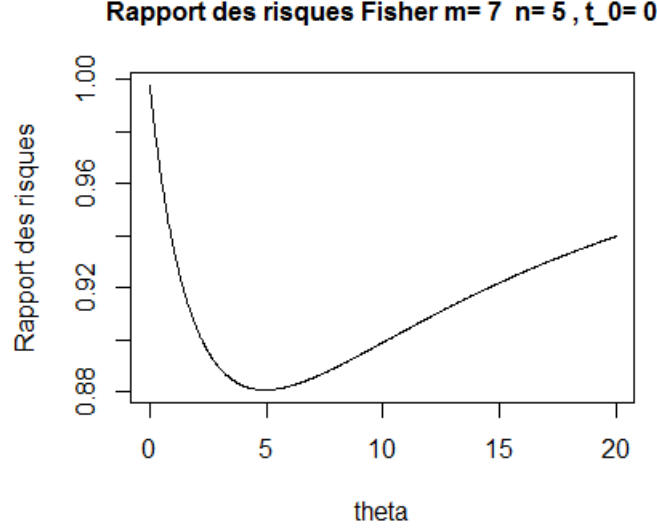


Figure 2.2 – Risques $\frac{R(\delta_{\psi^*, \theta})}{R(\delta_0, \theta)}$ d’une loi Fisher décentrée de degrés de liberté 7 et 5 puis de paramètre de décentralité $\theta \geq 0$.

On a obtenu au Lemme 1.1.10 que $\mathbb{E}_\theta(X) = \frac{m+1+n\theta}{n-m-1}$. On cherche des améliorations de l’ESB $\frac{(n-m-1)X-(m-1)}{n}$. Premièrement, on a les quantités suivantes :

$$h(k) = \mathbb{E}\left(\frac{n-m-1}{n}X - \frac{m-1}{n} \mid K = k\right) = \frac{2k}{n}$$

$$g_k(x) = \int_0^x \left[\frac{2k}{n} - \frac{(n-m-1)t - (m-1)}{n} \right] f_k(t) dt = \frac{2}{n} x(1+x) f_k(x).$$

En appliquant le Corollaire 2.1.13, on a que $\delta_\psi(X) = \frac{n-m-1}{n}X - \frac{m-1}{n} - \psi(X)$ avec ψ satisfaisant aux conditions (a), (b) domine l’ESB si

$$\psi(x) \geq \frac{-g_0(x)}{\int_0^x f_0(t) dt} = -\frac{2 x^{\frac{m-1}{2}} (1+x)^{-(\frac{n}{2}-1)}}{n B_{\frac{x}{1+x}}(\frac{m-1}{2}, \frac{n-m+1}{2})} = \psi^*(w), \quad \forall x > 0.$$

Si on considère $\psi(w) \equiv \psi^*(x)$, on remarque que $\lim_{x \rightarrow \infty} \psi^*(x) = 0$. Lorsque $m = 4$ et $n = 8$, on obtient le graphique du rapport des risques $\frac{R(\delta_{\psi^*, \theta})}{R(\delta_0, \theta)}$ à la figure 2.3. On remarque que le gain maximal se produit lorsque $\theta \approx 0,285$, c’est-à-dire lorsque $\rho^2 \approx 0,222$, avec une amélioration d’environ 12%.

2.2. POUR UN ESPACE PARAMÉTRIQUE BORNÉ INFÉRIEUREMENT

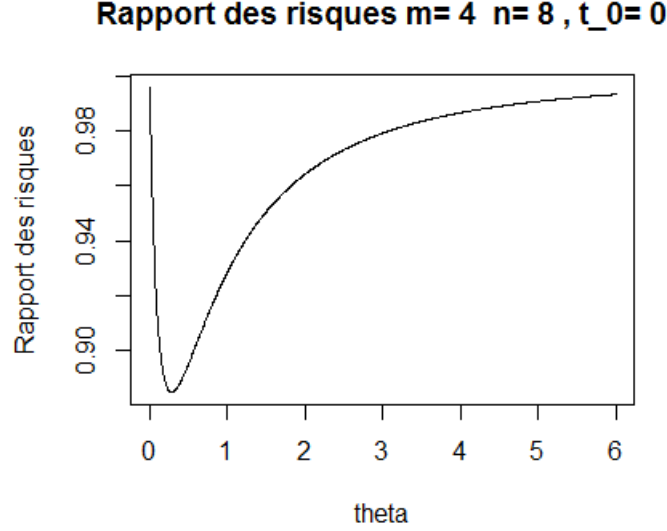


Figure 2.3 – Rapport des risques $\frac{R(\delta_{\psi^*}, \theta)}{R(\delta_0, \theta)}$ pour l'estimation de $\frac{\rho^2}{1-\rho^2}$ pour un échantillon de taille 8, un problème à 4 variables et $\theta \geq 0$.

2.2 Pour un espace paramétrique borné inférieurement

En présence d'une contrainte paramétrique $\theta \geq \theta_0$ avec $\theta_0 > 0$, plusieurs idées de la section précédente peuvent être réutilisées. En effet, la démarche menant au Corollaire 2.1.13 peut facilement être adaptée pour le problème d'estimation avec la contrainte $\theta \geq \theta_0$ afin d'obtenir un résultat similaire.

Théorème 2.2.1 (Théorème 3.1 de [23]). *Soit W issu du modèle (2.1) avec p_θ ayant un RVM croissant en K avec paramètre $\theta \geq \theta_0$. Supposons $\delta_0(W)$ strictement croissant en W et sans biais pour θ tel que $\frac{g_{k+1}(w)}{g_k(w)} \geq \frac{f_{k+1}(w)}{f_k(w)}$ pour tous $w > 0$, $k \in \mathbb{K}$. Posons $\delta_\psi(W) = \delta_0(W) - \psi(W)$ satisfaisant aux conditions (a) et (b). Sous la perte quadratique, δ_ψ domine δ_0 lorsque*

$$\psi(w) \geq \mathbb{E}_{\theta_0}(\delta_0(W) \mid W \leq w) - \mathbb{E}_{w, \theta_0}^L(h(L)), \quad \forall w > 0 \quad (2.14)$$

2.2. POUR UN ESPACE PARAMÉTRIQUE BORNÉ INFÉRIEUREMENT

si $\mathbb{E}_\theta(\delta_0^2(W)), \mathbb{E}_\theta(\psi^2(W)) < \infty$, où

$$L \sim F_l(w)p_{\theta_0}(l)/\mathbb{E}_{\theta_0}^K(F_K(w)). \quad (2.15)$$

Démonstration. On reprend la démonstration du Corollaire 2.1.13 avec $\theta \geq \theta_0$. \square

On remarque qu'on peut borner supérieurement $\mathbb{E}_{w,\theta_0}^L(h(L))$ par θ_0 et on trouve

$$\mathbb{E}_{\theta_0}(\delta_0(W) \mid W \leq w) - \mathbb{E}_{w,\theta_0}^L(h(L)) \geq \mathbb{E}_{\theta_0}(\delta_0(W) - \theta_0 \mid W \leq w)$$

où le terme de droite n'est pas sans rappeler la borne inférieure déduite de la méthode de Kubokawa au Théorème 1.4.10.

Pour illustrer ce résultat, on reprend les exemples de la Fisher décentrée et du mélange de lois inverse-gamma et gamma considérés à la sous-section précédente, mais avec une borne inférieure $\theta_0 > 0$ de l'espace paramétrique. On remarquera dans ces exemples que $\mathbb{E}_{w,\theta_0}^L(h(L))$ est proportionnel à un rapport de fonctions de répartition.

Exemple 2.2.2 (Suite de l'Exemple 2.1.11). *En appliquant le Théorème 2.2.1, une condition suffisante pour que $\delta_\psi(W) = \delta_0(W) - \psi(W)$ avec ψ satisfaisant aux conditions (a), (b) domine $\delta_0(W) = \frac{(l-1)W-m}{\alpha}$ sous la perte quadratique est que*

$$\psi(w) \geq \mathbb{E}_{\theta_0} \left(\frac{(l-1)W-m}{\alpha} \mid W \leq w \right) - \mathbb{E}_{w,\theta_0}^L(h(L)), \quad \forall x > 0.$$

De plus, on peut directement calculer $\mathbb{E}_{w,\theta_0}^L(h(L))$ comme $h(k) = \frac{k}{\alpha}$. On trouve

$$\mathbb{E}_{w,\theta}^L(L/\alpha) = \theta_0 \frac{F(w; \theta_0, l, m, \alpha + 1)}{F(w; \theta_0, l, m, \alpha)}$$

où $F(\cdot; \theta_0, l, m, \alpha)$ est la fonction de répartition de W pour les paramètres l, m, α et θ_0 .

Lorsqu'on considère $\psi^*(w) = \mathbb{E}_{\theta_0} \left(\frac{(l-1)W-m}{\alpha} \mid W \leq w \right) - \theta_0 \frac{F(w; \theta_0, l, m, \alpha + 1)}{F(w; \theta_0, l, m, \alpha)}$, on remarque que ψ^* est croissant en w puisque l'espérance tronquée l'est et $\mathbb{E}_{w,\theta_0}^L(h(L))$ est décroissant en w . De plus, on a que $\lim_{w \rightarrow \infty} \psi^*(w) = 0$. Lorsque $l = m = n = 3$ et $\theta_0 = 3$, on obtient le graphique des risques de δ_{ψ^*} et δ_0 à la figure 2.4. On remarque

2.2. POUR UN ESPACE PARAMÉTRIQUE BORNÉ INFÉRIEUREMENT

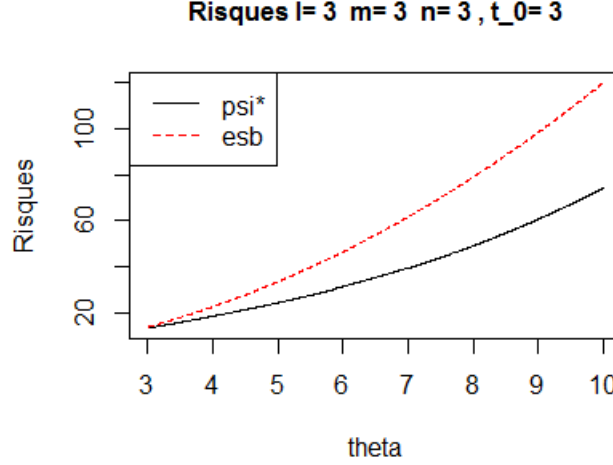


Figure 2.4 – Risques pour un mélange d'inverse-gamma et de gamma de paramètre d'échelle $\theta \geq 3$.

que le gain maximal se produit lorsque $\theta \approx 9,27$ avec une amélioration d'environ 38%, environ trois fois plus important que dans le cas non borné.

Exemple 2.2.3 (Suite de l'Exemple 2.1.14). On peut calculer directement le terme $\mathbb{E}_{w,\theta_0}^L(h(L))$ comme $h(k) = 2k$. On trouve

$$\mathbb{E}_{w,\theta_0}^L(2L) = \frac{\sum_{l \geq 0} 2l \frac{(\theta_0/2)^l e^{-\theta_0/2}}{l!} \frac{B_{\frac{mw/n}{1+mw/n}}(l+m/2, n/2)}{B(l+m/2, n/2)}}{\sum_{l \geq 0} \frac{(\theta_0/2)^l e^{-\theta_0/2}}{l!} \frac{B_{\frac{mw/n}{1+mw/n}}(l+m/2, n/2)}{B(l+m/2, n/2)}} = \theta_0 \frac{\mathbb{P}(F_{m+2,n}(\theta_0) \leq w)}{\mathbb{P}(F_{m,n}(\theta_0) \leq w)}.$$

Lorsqu'on considère $\psi^*(w) = \mathbb{E}_{\theta_0}(\frac{m(n-2)}{n}W - m \mid W \leq w) - \theta_0 \frac{\mathbb{P}(F_{m+2,n}(\theta_0) \leq w)}{\mathbb{P}(F_{m,n}(\theta_0) \leq w)}$, on remarque que ψ^* est croissant en w en vertu des propriétés des RVM et $\mathbb{E}_{w,\theta_0}^L(h(L))$ est décroissant en w . De plus, on a que $\lim_{w \rightarrow \infty} \psi^*(w) = 0$. Lorsque $m = 7$, $n = 5$ et $\theta_0 = 5$, on obtient le graphique du rapport des risques $\frac{R(\delta_{\psi^*}, \theta)}{R(\delta_0, \theta)}$ à la figure 2.5. On remarque que le gain maximal se produit lorsque $\theta \approx 13,544$ avec une amélioration d'environ 11,5%, un gain du même ordre que dans le cas non borné.

2.2. POUR UN ESPACE PARAMÉTRIQUE BORNÉ INFÉRIEUREMENT

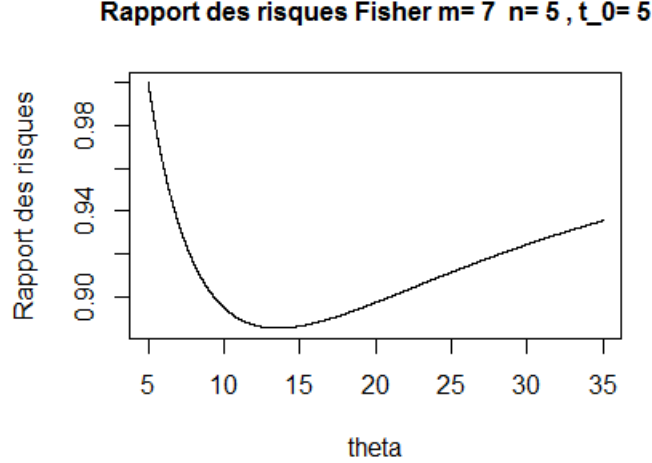


Figure 2.5 – Rapport des risques $\frac{R(\delta_{\psi^*}, \theta)}{R(\delta_0, \theta)}$ d’une Fisher décentrée de degrés de liberté 7 et 5 et de paramètre de décentralité $\theta \geq 5$.

Dans ce chapitre, on a présenté des résultats de dominance originaux par rapport à un estimateur cible δ_0 sans biais pour un paramètre θ , mais prenant des valeurs à l’extérieur de l’espace paramétrique. Ceux-ci, aussi applicables pour un espace paramétrique borné inférieurement, mènent à des estimateurs lisses prenant, sous certaines conditions, exclusivement des valeurs dans Θ . De plus, les théorèmes 2.1.6 et 2.2.1 reposent sur peu d’hypothèses :

- la loi conditionnelle de la densité de mélange a un RVM croissant ;
- l’estimateur sans biais est une fonction strictement croissante.

Divers exemples de densités de mélange où ces résultats s’appliquent ont été donnés tels que des mélanges dénombrables de lois gamma et bêta de type II (bêta, chi-deux et Fisher décentrées, distribution de $R^2/(1 - R^2)$ en régression multiple) ainsi que des modèles composites avec un seuil aléatoire tel que décrit par Pigeon et Denuit [40]. Ce dernier exemple pave la voie à beaucoup d’applications où la modélisation des données est difficile telles qu’en séquençage moléculaire où le paramètre d’échelle varie en fonction de l’emplacement [33] et la modélisation d’arrière-plan en segmentation d’images [41].

Pour compléter les travaux présentés dans ce chapitre, une exploration des liens

2.2. POUR UN ESPACE PARAMÉTRIQUE BORNÉ INFÉRIEUREMENT

avec les procédures bayésiennes serait souhaitable, une tâche qui semble toutefois laborieuse. En effet, pour $W \sim \chi_p^2(\theta)$ avec $\theta \geq 0$, [23] ont noté que l'estimateur de Bayes associé à une loi *a priori* gamma donné par Saxena et Alam [46] ne satisfait pas aux conditions du Corollaire 2.1.10. Une étude plus approfondie de cette loi est effectuée au Chapitre 3. On observera notamment que l'estimateur $(W - p)_+$ performe mieux que δ_{ψ^*} pour des petites valeurs de θ .

Chapitre 3

Sur l'estimation du paramètre de décentralité d'une loi chi-deux

La loi chi-deux décentrée est une distribution ayant de nombreuses applications liées aux tests d'hypothèses [15]. En effet, lorsqu'on considère un échantillon X_1, \dots, X_n indépendants tels que $X_i \sim \mathcal{N}(\mu_i, 1)$ pour $i \in \{1, \dots, n\}$, la statistique

$$W \stackrel{d}{=} \sum_{i=1}^n X_i^2 \quad (3.1)$$

est distribuée selon cette loi de probabilité avec paramètre de décentralité $\lambda = \sum_{i=1}^n \mu_i^2$. De plus, en appliquant la racine carrée à l'équation (3.1), on obtient la distribution de Rice qui compte plusieurs applications en analyse d'images [12], [48] et en traitement de signal [42], [19].

Considérons W un résumé stochastique de loi $\chi_p^2(\lambda)$ avec $p > 0$ connu et $\lambda \geq 0$ qu'on cherche à estimer. Pour le problème sans contrainte sur λ , deux estimateurs plausibles sont l'estimateur de vraisemblance maximale (EVM) $\hat{\lambda}_{\text{EVM}}$ ainsi que la partie positive de l'estimateur sans biais $\hat{\lambda}_+(W) = (W - p)_+ = \max\{W - p, 0\}$. Saxena et Alam [46] ont démontré la dominance stricte de $\hat{\lambda}_+$ sur $\hat{\lambda}_{\text{EVM}}$ sous la perte quadratique pour $p \geq 1$, soit lorsque

$$\mathbb{E}_\lambda [(\hat{\lambda}_+ - \lambda)^2] < \mathbb{E}_\lambda [(\hat{\lambda}_{\text{EVM}} - \lambda)^2] \quad \forall \lambda \geq 0. \quad (3.2)$$

3.1. SUR LES PROPRIÉTÉS DE L'EVM

Dans ce chapitre, on reprend les développements menant à ce résultat en plus de les généraliser pour les problèmes avec une contrainte de la forme $\lambda \geq \lambda_0$ avec $\lambda_0 > 0$. En procédant de la sorte, on propose des simplifications à l'aide d'inégalités de rapports de fonctions de Bessel modifiées du premier type démontrées par Amos [1]. On conclue ce chapitre avec les analyses bayésienne et prédictive pour un *a priori* gamma où l'on retrouve de jolies représentations de mélange telles que présentées par L'Moudden [28].

3.1 Sur les propriétés de l'EVM

La fonction de densité de $W \sim \chi_p^2(\lambda)$ pour $p > 0$ et $\lambda \geq 0$ admet la représentation suivante :

$$f_\lambda(w) = \frac{1}{2} e^{-\frac{w+\lambda}{2}} \left(\frac{w}{\lambda} \right)^{\frac{p/2-1}{2}} I_{p/2-1}(\sqrt{\lambda w}) \quad (3.3)$$

avec $I_\nu(\cdot)$ la fonction de Bessel modifiée du premier type d'ordre ν donnée par

$$I_\nu(x) = (x/2)^\nu \sum_{k \geq 0} \frac{(x/2)^{2k}}{k! \Gamma(\nu + k + 1)} \quad (3.4)$$

lorsque $x \geq 0$. En raison de cette structure, l'estimateur $\hat{\lambda}_{\text{EVM}}$ n'est pas accessible sous une forme close. Néanmoins, Saxena et Alam [46] ont bien saisi certaines propriétés clés de $\hat{\lambda}_{\text{EVM}}$ comme fonction de W et nécessaires pour leur démonstration de l'inégalité (3.2) où le rapport

$$Q_\nu(w) = \frac{I_\nu(w)}{I_{\nu-1}(w)}$$

pour $w \geq 0$ et $\nu \geq 1/2$ joue un rôle déterminant. Ce quotient et ses propriétés ont notamment servi pour la représentation d'estimateurs bayésiens ainsi que pour l'évaluation du risque fréquentiste sous la perte quadratique pour l'estimation de la moyenne μ d'une loi normale multivariée $\mathcal{N}_p(\mu, \sigma^2 1_p)$ avec contrainte $\|\mu\| \leq m$ [43], [3], [55], [30], [9].

Lemme 3.1.1. *Le rapport $Q_\nu(w)$ est croissant et concave en w avec $\lim_{w \rightarrow 0^+} Q_\nu(w) =$*

3.1. SUR LES PROPRIÉTÉS DE L'EVM

0 et $\lim_{w \rightarrow \infty} Q_\nu(w) = 1$. De plus, pour $\nu \geq 1$ et $w > 0$, on trouve les bornes suivantes :

$$\frac{w}{\nu - 1/2 + \sqrt{w^2 + (\nu + 1/2)^2}} \leq Q_\nu(w) \leq \frac{w}{\nu - 1/2 + \sqrt{w^2 + (\nu - 1/2)^2}}. \quad (3.5)$$

Enfin, $\frac{Q_\nu(w)}{w}$ est décroissant en w avec $\lim_{w \rightarrow 0^+} \frac{Q_\nu(w)}{w} = \frac{1}{2\nu}$ et $\lim_{w \rightarrow \infty} \frac{Q_\nu(w)}{w} = 0$.

Démonstration. Les inégalités (3.5) proviennent de l'article d'Amos [1] et les autres propriétés sont tirées du livre de Watson [52]. \square

À l'aide de ces propriétés, on trouve les prochains résultats décrivant $\hat{\lambda}_{EVM}$.

Lemme 3.1.2. Posons $g_w(\lambda) = \sqrt{\frac{w}{\lambda}} Q_{p/2}(\sqrt{w\lambda})$ avec $w, \lambda > 0$ et $g_w^{-1}(\lambda)$ son inverse pour un w fixe.

i) Pour $W \sim \chi_p^2(\lambda)$ avec $\lambda \geq 0$, on a que

$$\hat{\lambda}_{EVM}(w) = \begin{cases} 0 & \text{si } w \leq p \\ g_w^{-1}(1) & \text{sinon.} \end{cases}$$

ii) Pour $W \sim \chi_p^2(\lambda)$ avec $\lambda \geq \lambda_0 > 0$, on a que

$$\hat{\lambda}_{EVM}(w) = \begin{cases} \lambda_0 & \text{si } w \leq w_1 \\ g_w^{-1}(1) & \text{sinon} \end{cases}$$

où $w_1 \in [p, \infty)$ est la solution en w de $g_w(\lambda_0) = 1$.

Démonstration. Lorsqu'on dérive la vraisemblance donnée à l'équation (3.3), on obtient

$$\frac{\partial}{\partial \lambda} \log f_\lambda(w) = -\frac{1}{2} - \left(\frac{p/2 - 1}{2\lambda} \right) + \frac{1}{2} \sqrt{\frac{w}{\lambda}} \frac{I'_{\frac{p}{2}-1}(\sqrt{w\lambda})}{I_{\frac{p}{2}-1}(\sqrt{w\lambda})} = \frac{1}{2} \left(\frac{w}{\sqrt{w\lambda}} Q_{\frac{p}{2}}(\sqrt{w\lambda}) - 1 \right),$$

en utilisant la propriété $\frac{d}{dx} I_\nu(x) = I_{\nu+1}(x) + \frac{\nu}{x} I_\nu(x)$ qui peut être vérifiée à partir de

3.1. SUR LES PROPRIÉTÉS DE L'EVM

l'équation (3.4). Puisque $\lim_{\lambda \rightarrow \infty} g_w(\lambda) = 0$ en vertu du Lemme 3.1.1, on obtient que

$$\arg \max_{\lambda \geq \lambda_0} f_\lambda(w) = \begin{cases} \lambda_0 & \text{si } g_w(\lambda) \leq 1 \\ g_w^{-1}(1) & \text{sinon.} \end{cases}$$

Enfin, l'évaluation $\lim_{\lambda \rightarrow 0^+} g_w(\lambda) = w/p$ à l'aide du Lemme 3.1.1 mène à i) et établit que $w_1 \geq p$, ce qui complète la démonstration. \square

De plus, on trouve que l'EVM a les propriétés suivantes.

Lemme 3.1.3. *Pour $W \sim \chi_p^2(\lambda)$ avec $\lambda \geq 0$ et $p \geq 2$, on a que*

- i) $\hat{\lambda}_{EVM}(w)$ est croissant en w ;
- ii) pour $w \geq p$,

$$\underline{\hat{\lambda}} = w - p + 1 - \frac{p}{w} \leq \hat{\lambda}_{EVM}(w) \leq w - p + 1 = \bar{\hat{\lambda}} \quad (3.6)$$

- iii) $\hat{\lambda}_{EVM}(w) - w$ est strictement croissant en w pour $w \geq p$.

Remarque 3.1.4. *La propriété i) est issue du fait que f_λ est log-concave pour $p \geq 2$ en plus d'avoir un RVM croissant en W avec paramètre λ . En effet, comme l'ont noté Lehmann et Romano [26], cette dernière propriété est équivalente à $\frac{\partial^2}{\partial \lambda \partial w} \log f_\lambda(w) \geq 0$. On trouve (3.6) à l'aide des inégalités données par Amos [1], présentées à l'équation (3.5), et iii) est dû à Saxena et Alam [46]. Au lieu d'obtenir (3.6), ces derniers ont plutôt démontré, pour $p \geq 1$, que*

$$w - p \leq \hat{\lambda}_{EVM} \leq w - p + 1, \quad (3.7)$$

obtenant ainsi une borne inférieure moins précise. Enfin, le Lemme 3.1.3 tient sous la contrainte $\lambda \geq \lambda_0$ avec $w \geq w_1$.

Démonstration du Lemme 3.1.3. Puisque $g_w(\lambda)$ est croissant en w et décroissant λ , on trouve que $\hat{\lambda}_{EVM}(w) = g_w^{-1}(1)$ est une fonction croissante en w . En vertu de l'expression (3.5), on trouve que

$$g_{1,w}(\lambda) \leq g_w(\lambda) \leq g_{2,w}(\lambda)$$

3.1. SUR LES PROPRIÉTÉS DE L'EVM

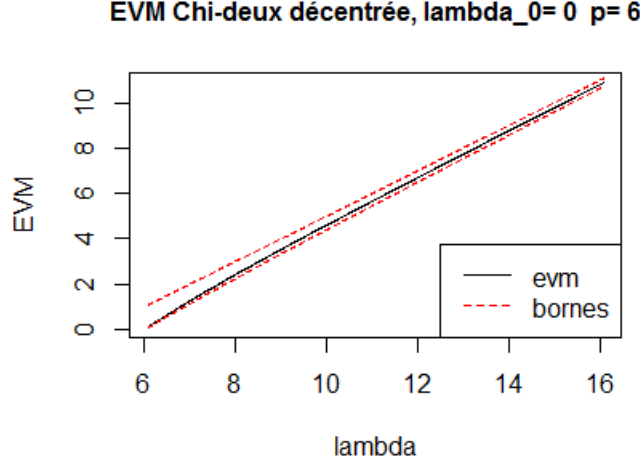


Figure 3.1 – Bornes de l'EVM pour une chi-deux décentrée de $p = 6$ degrés de liberté et de paramètre de décentralité $\lambda \geq 0$.

où $g_{1,w}(\lambda) = \frac{2w}{p-1+\sqrt{4w\lambda+(p+1)^2}}$ et $g_{2,w}(\lambda) = \frac{2w}{p-1+\sqrt{4w\lambda+(p-1)^2}}$. Ainsi, comme $\hat{\lambda}_{\text{EVM}}(w) = g_w^{-1}(1)$ pour $w \geq p$ et $g_w(\lambda)$ est décroissant en λ , on trouve que

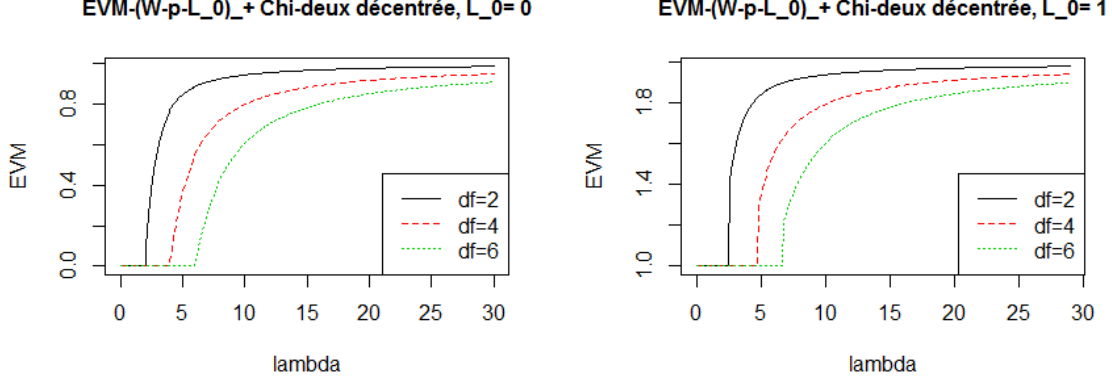
$$g_{1,w}^{-1}(1) \leq \hat{\lambda}_{\text{EVM}}(w) \leq g_{2,w}^{-1}(1), \quad \forall w \geq p.$$

Puisque $g_{1,w}(\hat{\lambda}) = g_{2,w}(\hat{\lambda}) = 1$, on obtient l'expression (3.6). On réfère à la démonstration de Saxena et Alam [46] pour iii) \square

On remarque que l'écart entre les bornes données à la propriété (3.6) du Lemme ci-dessus est décroissant en w , ce qu'on peut observer à la figure 3.1.

La condition iii du Lemme ci-dessus sera clé pour démontrer les résultats de dominance de la section suivante. Aussi, on illustre graphiquement la fonction $\hat{\lambda}_{\text{EVM}}(W) - (W - p - \lambda_0)_+$ à la figure 3.2. On note que celle-ci est non décroissante et non négative. De plus, on remarque que l'EVM est constant sur un intervalle légèrement plus grand à la figure 3.2b qu'à la figure 3.2a en raison de la borne inférieure de l'espace paramétrique.

3.2. COMPARAISONS DE RISQUES ET RÉSULTATS DE DOMINANCE



(a) Pour $\lambda_0 = 0$.

(b) Pour $\lambda_0 = 1$.

Figure 3.2 – La différence $\hat{\lambda}_{\text{EVM}}(W) - (W - p - \lambda_0)_+$ pour une $\chi_p^2(\lambda)$ avec $\lambda \geq \lambda_0$ et $p = 2, 4, 6$.

3.2 Comparaisons de risques et résultats de dominance

On reprend premièrement le résultat de dominance de $\hat{\lambda}_+$ sur $\hat{\lambda}_{\text{EVM}}$ démontré par Saxena et Alam [46]. À l'aide de la propriété iii) du Lemme 3.1.3, on peut décomposer $\hat{\lambda}_{\text{EVM}}$ comme la somme de la partie positive de l'ESB et d'une fonction croissante g . Cette approche demeure notamment valide pour démontrer la dominance pour le problème avec la contrainte inférieure.

Lemme 3.2.1. *Soient $X \in \mathbb{R}$ de densité $f(\cdot; \theta)$ absolument continue pour $\theta \in \mathbb{R}$ et δ_0 un estimateur de θ non décroissant de biais non négatif. Alors, pour $\delta_1(X) = \delta_0(X) + g(X)$ avec $g \not\equiv 0$ non négatif et non décroissant, δ_0 domine δ_1 sous la perte quadratique $(\delta - \theta)^2$.*

Démonstration. Comme δ_0 et g sont non décroissants, on a que $\text{Cov}(\delta_0, g) \geq 0$. D'où, $\text{Var}(\delta_1) = \text{Var}(\delta_0) + \text{Var}(g) + 2\text{Cov}(\delta_0, g) \geq \text{Var}(\delta_0)$. Puisque $\text{Var}(\delta_0) = R(\delta_0, \theta) - b_\theta^2(\delta_0)$ et que $\text{Var}(\delta_1) = R(\delta_1, \theta) - (b_\theta(\delta_0) + \mathbb{E}(g))^2$, on obtient que

$$R(\delta_0, \theta) \leq R(\delta_1, \theta) - (2b_\theta(\delta_0)\mathbb{E}(g) + \mathbb{E}(g)^2) \leq R(\delta_1, \theta). \quad \square$$

3.2. COMPARAISONS DE RISQUES ET RÉSULTATS DE DOMINANCE

Corollaire 3.2.2. *Soit $W \sim \chi_p^2(\lambda)$ avec $\lambda \geq \lambda_0 \geq 0$. Pour $p \geq 2$, $\hat{\lambda}_+$ domine $\hat{\lambda}_{EVM}$.*

Remarque 3.2.3. *Saxena et Alam [46] ont démontré le résultat pour $p \geq 1$ et $\lambda_0 = 0$.*

Démonstration du Corollaire 3.2.2. Considérons premièrement $\lambda_0 = 0$. En réutilisant la notation du Lemme 3.2.1, on pose

$$\begin{aligned}\delta_0(W) &= \hat{\lambda}_+(W) = (W - p)_+ \\ \delta_1(W) &= \hat{\lambda}_{EVM}(W) = (W - p + g_1(W))\mathbb{1}_{(p,\infty)}(W).\end{aligned}$$

Ainsi, on a que $\delta_1(W) = \delta_0(W) + g(W)$ où $g(W) = g_1(W)\mathbb{1}_{(p,\infty)}(W)$. De plus, en vertu du Lemme 3.1.3, on trouve que g est une fonction non décroissante et non négative. D'où, en appliquant le Lemme 3.2.1, on trouve que $\hat{\lambda}_+$ domine $\hat{\lambda}_{EVM}$ sous la perte quadratique.

De même, pour $\lambda_0 > 0$, on pose

$$\begin{aligned}\tilde{\delta}_0(W) &= \lambda_0 + (W - p - \lambda_0)_+ = \lambda_0 + (\delta_0 - \lambda_0)_+, \\ \tilde{\delta}_1(W) &= \lambda_0 + (\delta_1(W) - \lambda_0)_+.\end{aligned}$$

Ainsi, on a que $\tilde{\delta}_1(W) = \tilde{\delta}_0(W) + \tilde{g}(W)$ avec

$$\tilde{g}(W) = (\delta_1(W) - \lambda_0)_+ - (\delta_0(W) - \lambda_0)_+ = (\delta_0(W) + g(W) - \lambda_0)_+ - (\delta_0(W) - \lambda_0)_+.$$

Comme g est une fonction non décroissante et non négative, on trouve que \tilde{g} l'est aussi. D'où, en appliquant le Lemme 3.2.1, on trouve que $\hat{\lambda}_+$ domine $\hat{\lambda}_{EVM}$ sous la perte quadratique. \square

Il est assez étonnant que la simple projection de l'ESB sur l'espace paramétrique domine l'EVM. Dans le prochain exemple, on compare le rapport des risques $\frac{R(\hat{\lambda}_{EVM}, \lambda)}{R(\hat{\lambda}_+, \lambda)}$ pour plusieurs degrés de liberté p et bornes inférieures de l'espace paramétrique λ_0 .

Exemple 3.2.4. *Pour $\lambda_0 \in \{0, 2, 4, 6\}$, on trouve les rapports des risques $\frac{R(\hat{\lambda}_{EVM}, \lambda)}{R(\hat{\lambda}_+, \lambda)}$ à la figure 3.3. On remarque premièrement que, plus λ_0 est grand, moins le gain au niveau du rapport est important. En effet, pour une $\chi_2^2(\lambda)$, lorsque $\lambda_0 = 0$, le rapport maximal est supérieur à 1,4 alors qu'il est autour de 1,25 pour $\lambda_0 = 6$. De plus, l'effet*

3.2. COMPARAISONS DE RISQUES ET RÉSULTATS DE DOMINANCE

du degré de liberté s'amointrit avec la croissance de la borne inférieure comme l'écart entre les courbes est nettement plus prononcé lorsque $\lambda_0 = 0$ qu'en $\lambda_0 = 6$. Ces écarts sont vraisemblablement dûs à une question d'échelle, comme en témoigne la figure 3.4 où on illustre les risques quadratiques des estimateurs $W - p$, $\hat{\lambda}_{EVM}$ et $\hat{\lambda}_+$ pour $\lambda_0 \in \{0, 2\}$.

Bien que $\hat{\lambda}_+$ domine l'EVM, il n'en demeure pas moins que Chow [5] a démontré qu'il est inadmissible. Dans le Chapitre 2 de ce mémoire, on a étudié une classe d'estimateurs lisses pour le paramètre de décentralité λ de la forme $\hat{\lambda}_\psi(W) = W - \psi(W)$ dominant l'ESB pour les problèmes d'estimations avec ou sans contrainte paramétrique. Des estimateurs lisses compris dans l'espace paramétrique sont notamment inclus dans la classe d'estimateurs proposée, bien qu'on ne détermine pas s'ils sont des règles de Bayes et donc leur admissibilité. Bien qu'on ne cherche pas directement à dominer $\hat{\lambda}_+$, $W - p$ demeure une bonne cible puisque cet estimateur est minimax pour la perte quadratique pondérée, ce qu'on démontre à l'Exemple 3.3.4. De plus, cette approche permet de déduire une classe d'estimateurs qui comprend notamment $\hat{\lambda}_+$.

Lemme 3.2.5 (Exemples 2.1 et 3.1 de [23]). *Soit $W \sim \chi_p^2(\lambda)$ avec $\lambda \geq \lambda_0$. Alors, l'ESB $W - p$ est dominé par $\hat{\lambda}_\psi(W) = W - \psi(W)$ sous la perte quadratique lorsque*

- a) $\psi(w)$ est absolument continue, non décroissante en w et $\lim_{w \rightarrow \infty} \psi(w) = p$;
- b) pour $F_{p,\lambda}(\cdot)$ la fonction de répartition d'une $\chi_p^2(\lambda)$,

$$\psi(w) \geq \begin{cases} \int_0^w t^{p/2} e^{-t/2} dt / \int_0^w t^{p/2-1} e^{-t/2} dt & \text{si } \lambda_0 = 0 \\ \mathbb{E}_{\lambda_0}(W \mid W \leq w) - \lambda_0 \frac{F_{p+2,\lambda_0}(w)}{F_{p,\lambda_0}(w)} & \text{si } \lambda_0 > 0. \end{cases}$$

Exemple 3.2.6. *Pour $W \sim \chi_p^2(\lambda_0)$ avec $\lambda_0 \in \{0, 2\}$, on considère les estimateurs $\delta_0(W) = W - p$, $\hat{\lambda}_+$ et $\lambda_{\psi^*}(W) = W - \psi^*(W)$ avec*

$$\psi^*(w) = \begin{cases} \int_0^w t^{p/2} e^{-t/2} dt / \int_0^w t^{p/2-1} e^{-t/2} dt & \text{si } \lambda_0 = 0 \\ \mathbb{E}_{\lambda_0}(W \mid W \leq w) - \lambda_0 \frac{F_{p+2,\lambda_0}(w)}{F_{p,\lambda_0}(w)} & \text{si } \lambda_0 > 0. \end{cases}$$

On obtient leur risque fréquentiste respectif sous la perte quadratique à la figure 3.4.

3.2. COMPARAISONS DE RISQUES ET RÉSULTATS DE DOMINANCE

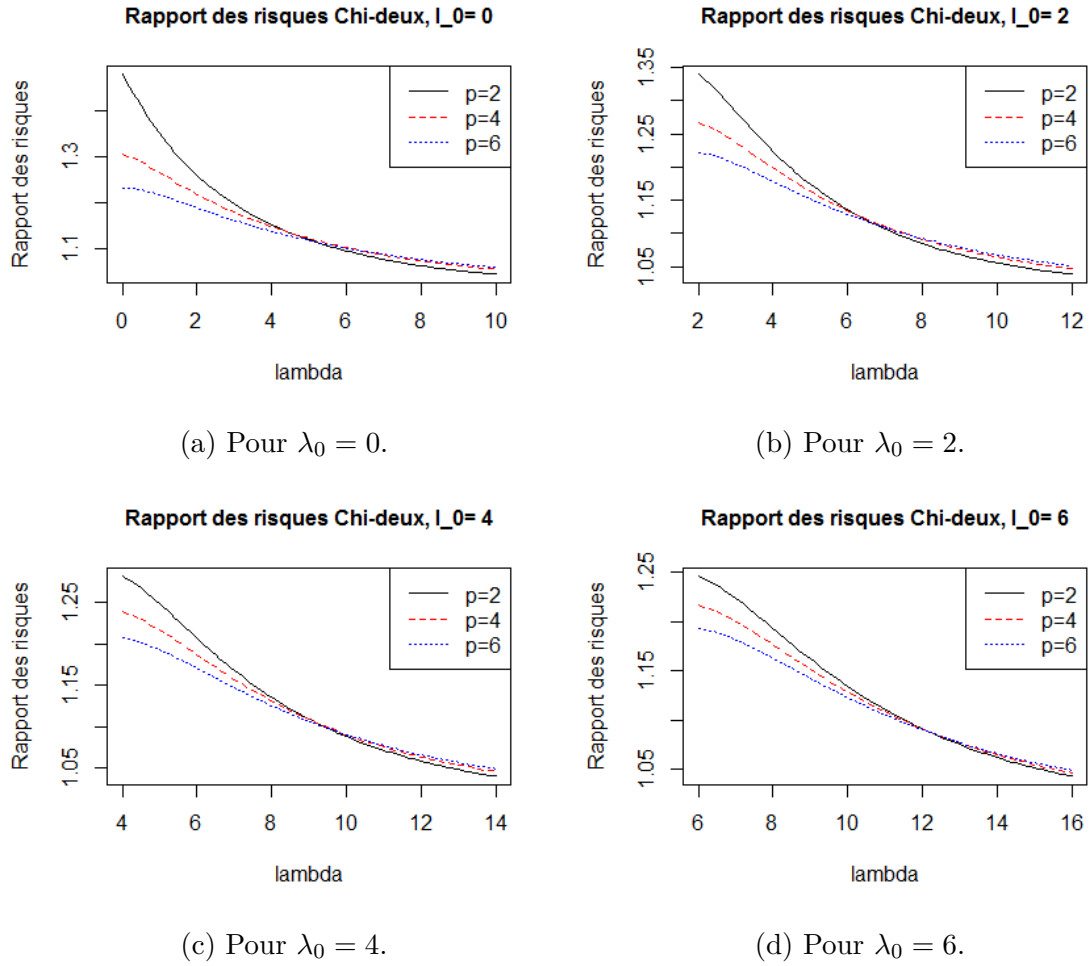


Figure 3.3 – Rapport des risques $\frac{R(\hat{\lambda}_{\text{EVM},\lambda})}{R(\hat{\lambda}_{+,\lambda})}$ d'une $\chi_p^2(\lambda)$ avec $\lambda \geq \lambda_0$ pour $p = 2, 4, 6$ et $\lambda_0 \in \{0, 2, 4, 6\}$.

3.3. INFÉRENCES BAYÉSIENNE ET PRÉDICTIVE

On note premièrement que les risques sont plus étalés lorsque $\lambda_0 = 2$. De plus, les risques de $\hat{\lambda}_{\psi^*}$ et l'ESB δ_0 semblent être égaux en λ_0 pour tout λ_0 . Lorsque $\lambda_0 = 0$, on peut démontrer cette propriété assez aisément. Enfin, on remarque que $\hat{\lambda}_+$ performe le mieux pour les petites valeurs de λ et $\hat{\lambda}_{\psi^*}$ pour les plus grandes. Lorsque la borne inférieure de l'espace paramétrique est précise, considérer $\hat{\lambda}_+$ semble donc être un bon choix. De plus, il est intéressant de noter que ces deux estimateurs appartiennent à la classe donnée par Kubokawa et coll. [23], présentée au Chapitre 2, en considérant

$$\hat{\lambda}_+(W) = \max\{\lambda_0, W - p\} = W - \psi(W)$$

où $\psi(t) = \min\{t - \lambda_0, p\}$.

3.3 Inférences bayésienne et prédictive

On peut aussi estimer le paramètre de décentralité avec l'approche bayésienne. Une approche avec des lois non informative est présentée par Berger et coll. [2]. On choisit ici un *a priori* gamma pour ses propriétés mathématiques. De plus, un modèle similaire a été étudié pour la distribution de Rice par Kim et coll. [17] où l'on s'intéresse au mode *a posteriori* plutôt qu'à la règle de Bayes.

En tirant profit de la représentation de mélange d'une chi-deux décentrée à p degrés de liberté et de paramètre de décentralité λ

$$\begin{aligned} W \mid K = k &\sim \mathcal{G}(k + p/2, 2) \\ K &\sim \mathcal{P}(\lambda/2), \end{aligned}$$

on obtient que les densités *a posteriori* et prédictive se décomposent elles aussi comme des mélanges du même type, une structure présentée par L'Moudden [28]. Pour obtenir celle-ci, il est nécessaire de recourir à la généralisation suivante.

Définition 3.3.1. On appelle ${}_pF_q(\boldsymbol{\alpha}; \boldsymbol{\beta}; x)$ la fonction hypergéométrique généralisée si elle admet la représentation

$${}_pF_q(\boldsymbol{\alpha}; \boldsymbol{\beta}; x) = \sum_{k \geq 0} \frac{(\boldsymbol{\alpha})_k}{(\boldsymbol{\beta})_k} \frac{x^k}{k!} \quad (3.8)$$

3.3. INFÉRENCES BAYÉSIENNE ET PRÉDICTIVE

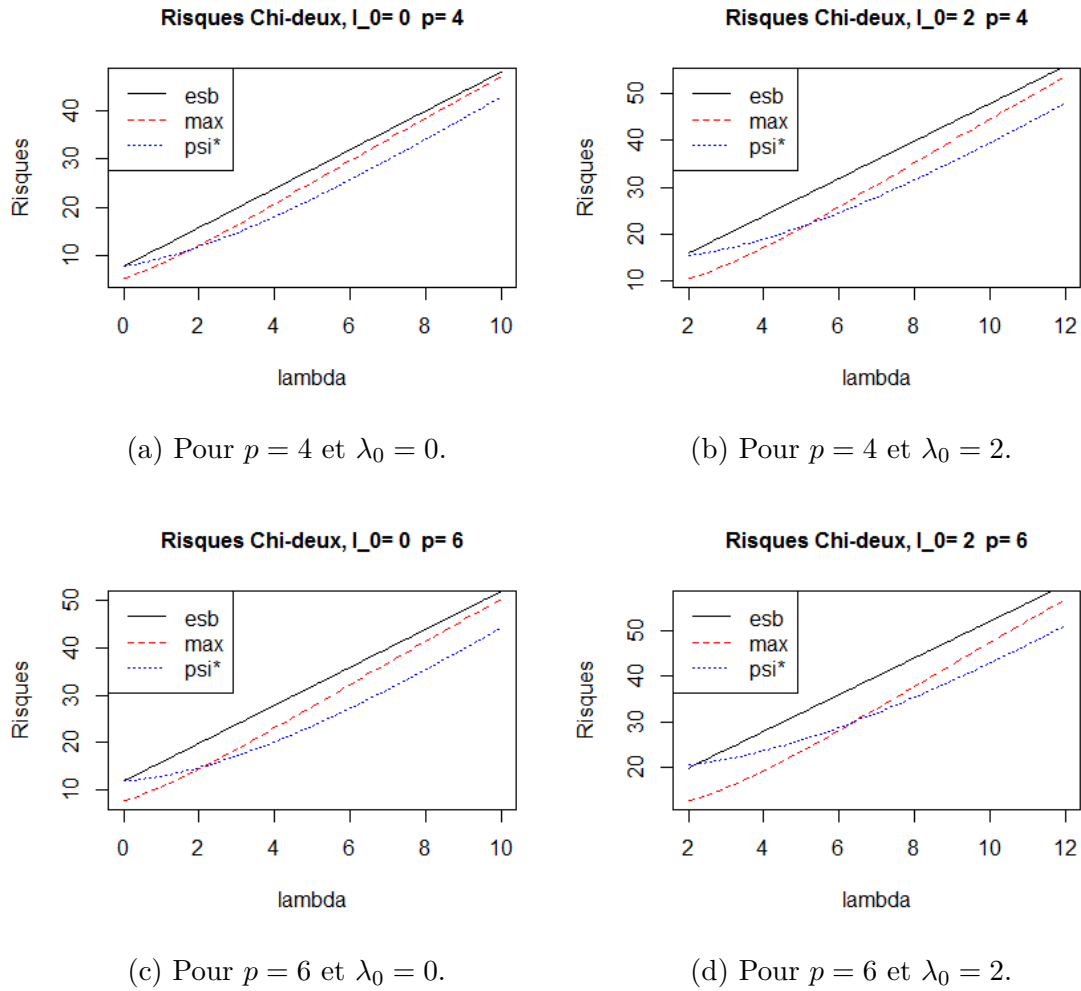


Figure 3.4 – Risques quadratiques de $W - p$, $\hat{\lambda}_+$ et $\hat{\lambda}_\psi$ pour $W \sim \chi_p^2(\lambda)$ avec $\lambda \geq \lambda_0$ et $p = 4, 6$.

3.3. INFÉRENCES BAYÉSIENNE ET PRÉDICTIVE

où α, β sont des vecteurs de dimensions respectives p et q et x un réel¹. Pour un vecteur réel $\mathbf{v} = (v_1, v_2, \dots, v_m)$ donné et un entier positif k , le symbole de Pochhammer est $(\mathbf{v})_k = \prod_{i=1}^m \prod_{j=0}^{k-1} (v_i + j)$ et $(v)_0 = 1$.

Lorsque $p = 0$, on a que ${}_0F_q(-; \beta; x) = \sum_{k \geq 0} \frac{1}{(\beta)_k} \frac{x^k}{k!}$ et, lorsque $q = 0$, on a que ${}_pF_0(\alpha; -; x) = \sum_{k \geq 0} (\alpha)_k \frac{x^k}{k!}$.

Définition 3.3.2. On dit que la variable discrète X est de loi hypergéométrique généralisée si sa fonction de masse p_λ est de la forme

$$p_\lambda(x) = \frac{(\alpha)_x \lambda^x}{(\beta)_x x!} \frac{1}{{}_pF_q(\alpha; \beta; \lambda)} \mathbb{1}_{\mathbb{N}}(x)$$

avec $\alpha \in \mathbb{R}_+^p$, $\beta \in \mathbb{R}_+^q$ et $\lambda \geq 0$ lorsque ${}_pF_q(\alpha; \beta; \lambda) < \infty$. On note $X \sim \text{Hyp}(\alpha; \beta; \lambda)$.

Cette distribution discrète est présentée plus en détail dans Johnson et coll. [13], où l'on discute notamment des conditions pour que ${}_pF_q(\alpha; \beta; \lambda)$ converge. Cette loi de probabilité apparaît aussi lorsqu'on effectue l'inférence bayésienne du coefficient de détermination comme présenté par Marchand [54].

Exemple 3.3.1. Soit $\lambda > 0$. Voici quelques cas particuliers :

1. lorsque $\alpha = \beta$, on trouve que $\text{Hyp}(\alpha; \alpha; \lambda) \stackrel{d}{=} \text{Hyp}(-; -; \lambda) \stackrel{d}{=} \mathcal{P}(\lambda)$;
2. lorsque $a > 0$ et $\lambda < 1$, $\text{Hyp}(a; -; \lambda) \stackrel{d}{=} \mathcal{BN}(a, 1 - \lambda)$.

On peut facilement démontrer le résultat suivant.

Lemme 3.3.2. Lorsque $X \sim \text{Hyp}(\alpha; \beta; \lambda)$, on trouve que la fonction génératrice des probabilités est

$$\mathbb{E}(t^X) = \frac{{}_pF_q(\alpha; \beta; t\lambda)}{{}_pF_q(\alpha; \beta; \lambda)}$$

et que l'espérance est

$$\mathbb{E}_\lambda(X) = \lambda \frac{\prod_{i=1}^p \alpha_i {}_pF_q(\alpha + \mathbf{1}_p; \beta + \mathbf{1}_q; \lambda)}{\prod_{i=1}^q \beta_i {}_pF_q(\alpha; \beta; \lambda)}$$

où $\mathbf{1}_n = [1, \dots, 1] \in \mathbb{R}^n$.

1. Lorsque $p = q = 1$, elle est dite confluyente et, lorsque $p = 2$ et $q = 1$, elle est dite de Gauss.

3.3. INFÉRENCES BAYÉSIENNE ET PRÉDICTIVE

Inférence bayésienne

On considère ici une loi *a priori* gamma, à l'instar de Saxena et Alam [46], une distribution assez versatile sur les réels positifs qui présente des propriétés mathématiques intéressantes. Tel que présenté par L'Moudden [28], on obtient que la loi *a posteriori* est le mélange d'une loi gamma et d'une loi hypergéométrique généralisée, qui n'est pas sans rappeler celui de départ. De plus, ces résultats peuvent être adaptés pour la distribution de Rice.

Théorème 3.3.3. *Supposons $W \mid \lambda \sim \chi_p^2(\lambda)$ et $\lambda \sim \mathcal{G}(a, 1/b)$ pour $a, b > 0$, alors la loi *a posteriori* est unimodale pour $a > 1$ et admet la représentation*

$$\begin{aligned}\lambda \mid W, K = k &\sim \mathcal{G}\left(a + k, \frac{1}{b + 1/2}\right) \\ K \mid W = w &\sim \text{Hyp}\left(a; \frac{p}{2}; \frac{w}{4b + 2}\right).\end{aligned}\tag{3.9}$$

L'espérance *a posteriori* est $\mathbb{E}(\lambda \mid W = w) = \frac{2a}{2b+1} + \frac{w}{(2b+1)^2} \frac{a}{p/2} \frac{{}_1F_1(1+a; 1+p/2, \frac{w}{2(2b+1)})}{{}_1F_1(a; p/2, \frac{w}{2(2b+1)})}$.

Démonstration. On a que la densité *a posteriori* s'écrit

$$\begin{aligned}\pi(\lambda \mid W = w) &\propto \sum_{k \geq 0} \frac{w^{\frac{p}{2}+k-1} e^{-\frac{w}{2}}}{\Gamma\left(\frac{p}{2} + k\right) 2^{\frac{p}{2}+k}} \frac{\left(\frac{\lambda}{2}\right)^k e^{-\frac{\lambda}{2}}}{k!} \frac{\lambda^{a-1} e^{-b\lambda} b^a}{\Gamma(a)} \\ &\propto \sum_{k \geq 0} \lambda^{a+k-1} e^{-\lambda(1/2+b)} \frac{(w/4)^k}{\Gamma\left(\frac{p}{2} + k\right) k!} \\ &\propto \sum_{k \geq 0} \frac{\lambda^{a+k-1} e^{-\lambda(1/2+b)} (1/2 + b)^{a+k}}{\Gamma(a+k)} \frac{\left(\frac{a}{2}\right)_k}{\left(\frac{p}{2}\right)_k} \left(\frac{w}{2(2b+1)}\right)^k / k!.\end{aligned}$$

On reconnaît que le premier terme est celui de la densité d'une $\mathcal{G}(a + k, 2/(2b + 1))$ et que le deuxième est celui de la fonction de masse d'une $\text{Hyp}\left(a; p/2; \frac{x}{2(2b+1)}\right)$, d'où le mélange 3.9.

Pour l'unimodalité, la densité *a posteriori*

$$\pi(\lambda \mid W = w) = \frac{\lambda^{a-1} e^{-\lambda(b+1/2)} (b + 1/2)^a}{\Gamma(a)} \frac{{}_0F_1\left(-; \frac{p}{2}; \frac{w\lambda}{4}\right)}{{}_1F_1\left(a; \frac{p}{2}; \frac{w}{2(2b+1)}\right)},$$

3.3. INFÉRENCES BAYÉSIENNE ET PRÉDICTIVE

se décompose comme le produit de $g(\lambda) = \frac{\lambda^{a-1}e^{-\lambda(b+1/2)}}{\Gamma(a)(b+1/2)^{-a}}$ et $h(\lambda) = \frac{{}_0F_1\left(-; \frac{p}{2}; \frac{w\lambda}{4}\right)}{{}_1F_1\left(a; \frac{p}{2}; \frac{w}{2(2b+1)}\right)}$. Alors, pour $a > 1$, on trouve que $\frac{\partial}{\partial \lambda}\pi(\lambda | W = w) = g'h + gh'$ où g' est positive sur $(0, \frac{a-1}{b+1/2})$ puis négative pour $\lambda > \frac{a-1}{b+1/2}$ et $h' \geq 0$ pour tout $\lambda > 0$. D'où, la densité *a posteriori* est unimodale pour $a > 1$.

Finalement, on obtient l'espérance *a posteriori* à l'aide du Lemme 3.3.2. \square

Exemple 3.3.4. *Considérons $W \sim \chi_p^2(\lambda)$ avec $\lambda \sim \mathcal{G}(p/2, 1/b)$ où $b > 0$. Tel que présenté par Saxena et Alam [46], on trouve que la densité *a posteriori* admet la représentation*

$$\begin{aligned} \lambda | W, K = k &\sim \mathcal{G}\left(p/2 + k, \frac{2}{2b+1}\right) \\ K | W = w &\sim \mathcal{P}\left(\frac{w}{4b+2}\right), \end{aligned} \tag{3.10}$$

en vertu du Théorème 3.3.3 et de l'Exemple 3.3.1.1. Ainsi, le mélange (3.10) est très similaire à celui d'une $\chi_p^2\left(\frac{w}{2b+1}\right)$. Pour ce modèle, on trouve que l'espérance *a posteriori* est $\frac{p}{2b+1} + \frac{w}{(2b+1)^2}$.

On illustre les densités *a priori* et *a posteriori* pour $p \in \{2, 4\}$ et $b \in \{0, 01; 0, 5; 1\}$ à la figure 3.5. On remarque que, lorsque les observations sont plus grandes, les densités *a posteriori* sont plus dispersées. De plus, on choisit une petite valeur de paramètre b comme on s'intéressera à la règle de Bayes associée à ce genre paramétrisation à l'Exemple 3.3.5. On note aussi que, lorsque $b = 0, 5$, la loi *a priori* considérée est une chi-deux centrée à p degrés de liberté.

Enfin, comme les densités *a posteriori* sont unimodales en vertu du Théorème 3.3.3, la construction d'intervalles bayésiens de probabilité *a posteriori* maximale (« *highest posteriori density* ») est assez simple.

On observe à la figure 3.4 que les risques quadratiques sont non bornés. En effet, pour l'ESB $W - p$, on a que

$$R(W - p, \lambda) = \text{Var}_\lambda(W) = 2p + 4\lambda.$$

On peut donc espérer déterminer des estimateurs minimax en recourant à une normalisation de la perte. Dans l'exemple suivant, on effectue un tel développement et

3.3. INFÉRENCES BAYÉSIENNE ET PRÉDICTIVE

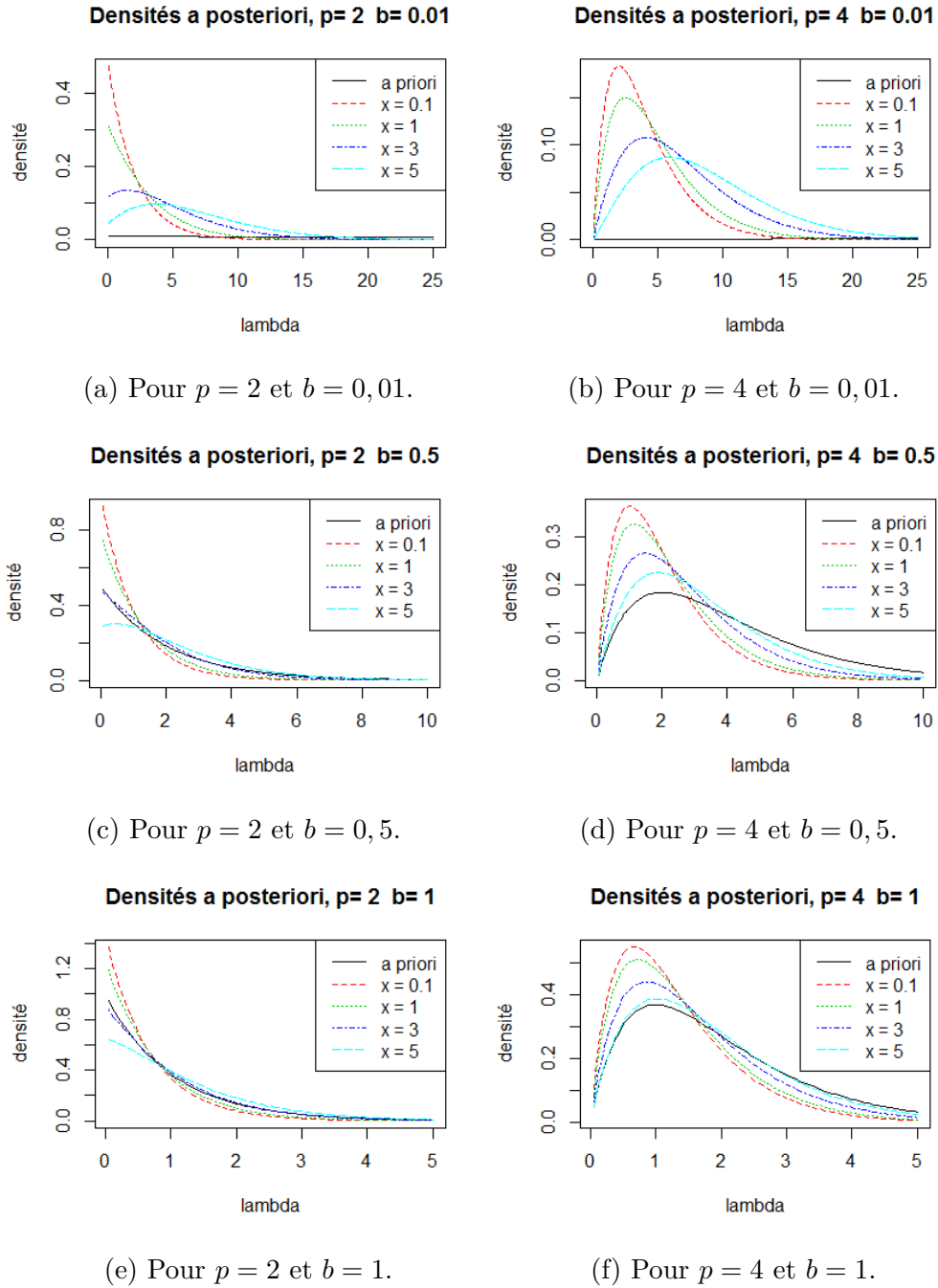


Figure 3.5 – Densités *a posteriori* pour une $\chi_p^2(\lambda)$ avec les observations $x \in \{0, 1; 1; 3; 5\}$, $p \in \{2, 4\}$ et $b \in \{0, 01; 0, 5; 1\}$.

3.3. INFÉRENCES BAYÉSIENNE ET PRÉDICTIVE

on identifie une règle minimax en suivant celui donné par [46].

Exemple 3.3.5 (Suite de l'Exemple 3.3.4). *Pour $\lambda \sim \mathcal{G}(p/2, 1/b)$, la règle de Bayes δ_π pour la perte quadratique est $\mathbb{E}(\lambda \mid W = w) = \frac{p}{1+2b} + \frac{w}{(1+2b)^2}$ et son risque associé est*

$$R(\delta_\pi, \lambda) = \text{Var}_\lambda \left[\frac{W}{(2b+1)^2} \right] + b_\lambda^2 \left[\frac{W}{(2b+1)^2} \right] = \frac{2p + 4\lambda + (2b+2)^2(p - 2b\lambda)^2}{(1+2b)^4}.$$

Ainsi, on obtient que le risque intégré de Bayes est

$$\begin{aligned} r(\pi, \delta_\pi) &= \int_{\mathbb{R}_+} R(\delta_\pi, \lambda) \pi(\lambda) d\lambda \\ &= \frac{16b^2p(b+1) + (3b+2)^2(2p^2b^3 + b^2(b-2)^2(p/2 + p^2/4))}{(2+b)^4}. \end{aligned}$$

De plus, en appliquant le Lemme 1.3.10, on trouve que $\delta_\pi(w) = \frac{p}{(1+2b)} + \frac{w}{(1+2b)^2}$ est la règle de Bayes associée à la loi a priori $\tilde{\pi}(\lambda) = \frac{\lambda+p/2}{\frac{1+b}{b}p/2} \frac{\lambda^{p/2-1}e^{-\frac{\lambda}{b}}}{\Gamma(p/2)b^{p/2}}$ et à la perte quadratique pondérée $\tilde{\rho}(\delta - \lambda) = \frac{(\delta-\lambda)^2}{\lambda+p/2}$. Dans ce cas, on a que le risque intégré de Bayes est

$$r(\tilde{\pi}, \delta_{\tilde{\pi}}) = \frac{r(\pi, \delta_\pi)}{\frac{b+1}{b}p/2} = \frac{4 + 16b(b+1)}{(2b+1)^4}.$$

Puisque $r(\tilde{\pi}, \delta_{\tilde{\pi}}) \rightarrow 4$ lorsque $b \rightarrow 0$ et que le risque de l'ESB est constant à 4, ce dernier ainsi que $\hat{\lambda}_+$ sont minimax pour la perte pondérée $\tilde{\rho}$ en vertu du Théorème 1.3.13. Tel que mentionné précédemment, Chow [5] a démontré que λ_+ demeure inadmissible. Trouver des estimateurs admissibles et minimax pour la chi-deux décentrée demeure à ce jour un problème ouvert [27].

On illustre les risques quadratiques pondérés de $W - p$ et $(W - p)_+$ ainsi que le risque minimax à la figure 3.6, qui semblent assez similaires entre les différents degrés de liberté considérés.

Analyse de densités prédictives

Un autre domaine important de l'analyse bayésienne est l'estimation de densités prédictives. Comme l'objectif est d'estimer la distribution d'une nouvelle observation sachant le modèle préalablement défini, la reproductibilité d'études scientifiques en

3.3. INFÉRENCES BAYÉSIENNE ET PRÉDICTIVE

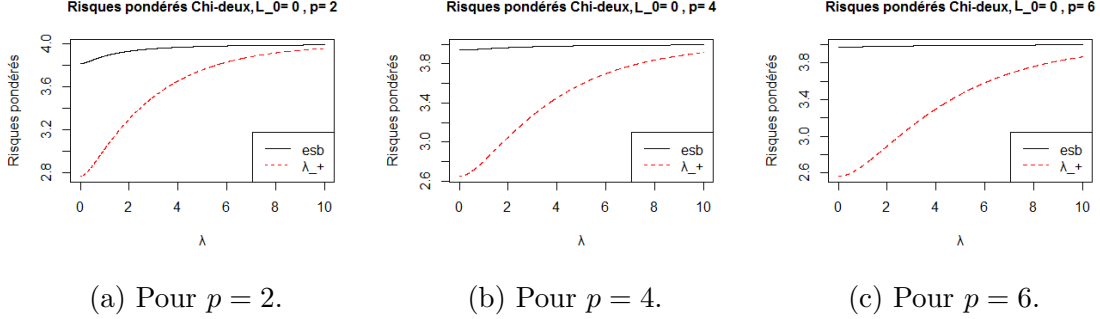


Figure 3.6 – Risques quadratiques pondérés de $W - p$ et $(W - p)_+$ pour $W \sim \chi_p^2(\lambda)$ avec $\lambda \geq 0$ et $p \in \{2, 4, 6\}$.

fait une application naturelle. À l’aide des résultats de L’Moudden [28], la densité prédictive obtenue dans le cadre du Théorème 3.3.6 demeure un mélange similaire à celui de la densité *a posteriori* donnée au Théorème 3.3.3. Cette représentation permet notamment d’évaluer directement la densité, sans recourir à des méthodes numériques basées sur la méthode de Monte-Carlo par chaînes de Markov (MCMC). Ces techniques, présentées en détails par Carlin et Louis [4], sont généralement nécessaires lorsque la densité prédictive n’a pas de forme close ou est trop difficile à évaluer.

De plus, ces résultats peuvent être adaptés pour la distribution de Rice, une loi de probabilité pour laquelle l’étude de densités prédictives ne semble pas être une pratique répandue.

Théorème 3.3.6. *Soient $X \mid \lambda \sim \chi_p^2(\lambda)$ et $Y \mid \lambda \sim \chi_q^2(\lambda)$ que l’on suppose conditionnellement indépendantes par rapport à λ . Pour $\lambda \sim \mathcal{G}(a, 1/b)$, on trouve que la densité prédictive de $Y \mid X$ admet la représentation*

$$\begin{aligned}
 Y \mid X, L, K = k &\sim \mathcal{G}\left(\frac{q}{2} + k, 2\right) \\
 K \mid X, L = l &\sim \mathcal{BN}\left(a + l, \frac{1}{2b + 2}\right) \\
 L \mid X = x &\sim \text{Hyp}\left(a; \frac{p}{2}; \frac{x}{4b + 4}\right).
 \end{aligned}$$

3.3. INFÉRENCES BAYÉSIENNE ET PRÉDICTIVE

Démonstration. On a vu au Théorème 3.3.3 que

$$\pi(\lambda \mid X = x) = \sum_{k \geq 0} \frac{\lambda^{a+k-1} e^{-\lambda(1/2+b)} (1/2+b)^{a+k} \frac{(a)_k}{(\frac{p}{2})_k} \left(\frac{x}{4b+2}\right)^k / k!}{\Gamma(a+k) {}_1F_1\left(a; \frac{p}{2}; \frac{x}{4b+2}\right)}.$$

On a donc que la densité prédictive s'écrit, pour $y > 0$,

$$\begin{aligned} f_{Y|X=x}(y) &= \int_{\Lambda} f_{\lambda}(y) \pi(\lambda \mid X = x) d\lambda \\ &\propto \sum_{k \geq 0} \sum_{l \geq 0} \frac{y^{\frac{q}{2}+k-1} e^{-\frac{y}{2}}}{\Gamma\left(\frac{q}{2}+k\right) 2^{\frac{q}{2}+k}} \frac{(a)_l \left(\frac{x}{4b+2}\right)^l}{\left(\frac{p}{2}\right)_l l!} \int_0^{\infty} \frac{\lambda^{a+l-1} e^{-\lambda(b+1)}}{\Gamma(a+l) (b+1/2)^{a+l} k!} d\lambda \\ &\propto \sum_{k \geq 0} \sum_{l \geq 0} \frac{y^{\frac{q}{2}+k-1} e^{-\frac{y}{2}}}{\Gamma\left(\frac{q}{2}+k\right) 2^{\frac{q}{2}+k}} \frac{\frac{(a+l)_k}{k!} \left(\frac{1}{2(b+1)}\right)^k}{{}_1F_0\left(a+l; -; \frac{1}{2(b+1)}\right)} \frac{\frac{(a)_l}{\left(\frac{p}{2}\right)_l} \left(\frac{x}{4(b+1)}\right)^l \frac{1}{l!}}{{}_1F_1\left(a; \frac{p}{2}; \frac{x}{4(b+1)}\right)}. \end{aligned}$$

Comme ${}_1F_0\left(a+l; -; \frac{1}{2(b+1)}\right) = \left(\frac{1}{2(b+1)}\right)^{-a-l}$, on obtient le résultat. \square

Exemple 3.3.7 (Suite de l'Exemple 3.3.4). *En considérant $a = p/2 = q/2$ au Théorème 3.3.6, on trouve que la densité a posteriori admet la jolie représentation*

$$\begin{aligned} Y \mid X, L, K = k &\sim \mathcal{G}\left(\frac{p}{2} + k, 2\right) \\ K \mid X, L = l &\sim \mathcal{BN}\left(p/2 + l, \frac{1}{2b+2}\right) \\ L \mid X = x &\sim \mathcal{P}\left(\frac{x}{4b+4}\right). \end{aligned}$$

À titre d'illustration, les densités prédictives lorsque $p = 4$ et $b \in \{0, 01; 0, 5; 1\}$ sont représentées à la figure 3.7.

3.3. INFÉRENCES BAYÉSIENNE ET PRÉDICTIVE

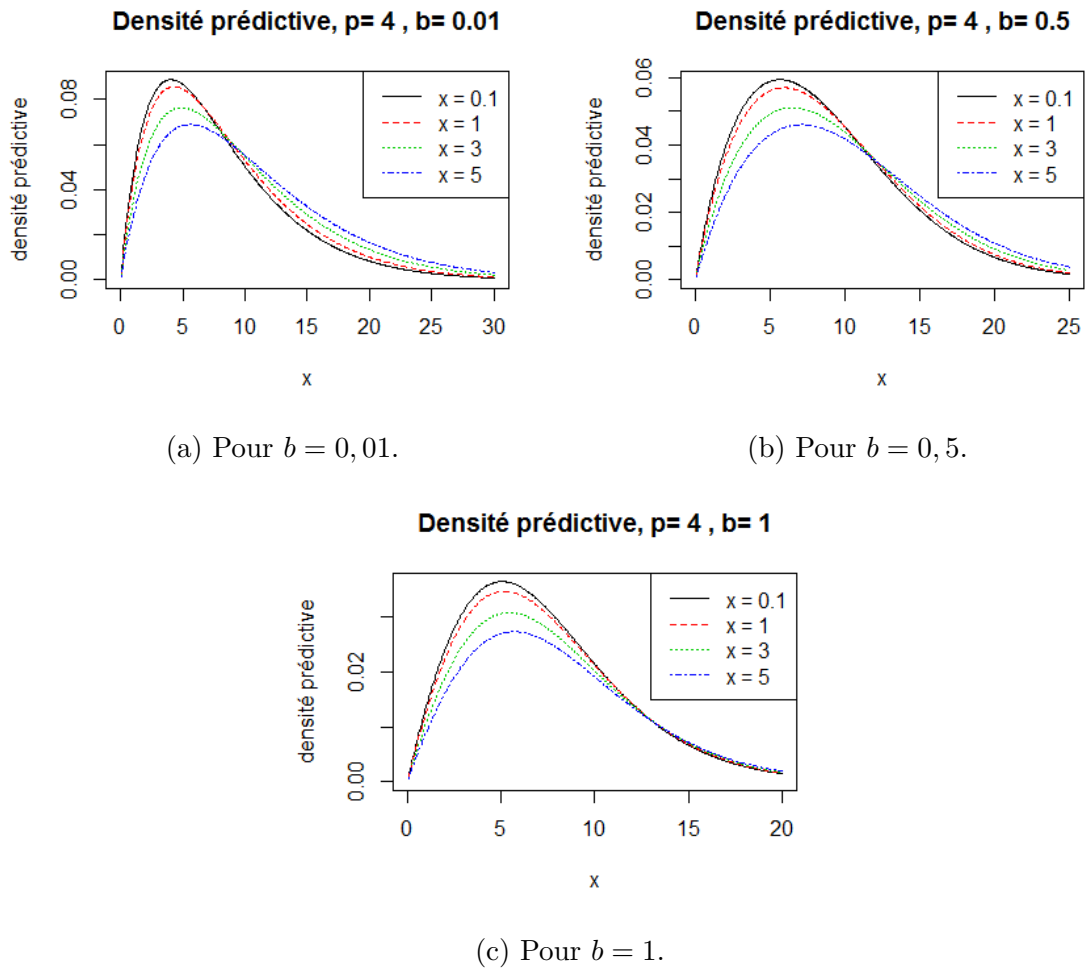


Figure 3.7 – Densités prédictives d'une $\chi_4^2(\lambda)$ pour les observations $x \in \{0, 1; 1; 3; 5\}$ et $b \in \{0,01; 0,5; 1\}$.

Conclusion

Tout au long de ce mémoire, on a traité de l'estimation ponctuelle pour des mélanges de lois avec une attention particulière pour la perte quadratique et effectué des comparaisons de fonctions de risque. Plusieurs résultats de dominance ont été obtenus, s'appliquant autant à des mélanges dénombrables qu'indénombrables, et des illustrations variées ont été données afin de témoigner de la généralité des résultats.

En effet, on a premièrement proposé une classe d'estimateurs lisses et appartenant à l'espace paramétrique qui dominant l'estimateur sans biais (ESB). De plus, ces résultats originaux qui généralisent ceux de Kubokawa et coll. [23], pouvant aussi être appliqués en présence d'une borne inférieure de l'espace paramétrique, requièrent peu d'hypothèses : la loi conditionnelle du modèle hiérarchique a un rapport de vraisemblance monotone (RVM) croissant et l'ESB du paramètre θ est une fonction croissante. Parmi les exemples présentés, on retrouve la loi de Fisher décentrée et le coefficient de détermination, un mélange de lois gamma et un modèle composite à seuil aléatoire inspiré de celui proposé par Pigeon et Denuit [40]. Ce dernier ouvre la porte à une grande classe d'applications où la modélisation des données est difficile, et ce, même en ayant recours aux mélanges dénombrables telles que le séquençage moléculaire [33] et la modélisation d'arrière-plan [41].

Par la suite, on a étudié plus en détail le problème d'estimation du paramètre de décentralité de la loi chi-deux en reprenant quelques résultats présentés par Saxena et Alam [46]. En comparant les fonctions de risque de quelques estimateurs, on a remarqué que la troncature de l'ESB sur l'espace paramétrique, qui domine l'estimateur de vraisemblance maximale (EVM), performe le mieux pour les petites valeurs que l'enveloppe inférieure de la classe donnée au Chapitre 2. Ainsi, lorsque la borne inférieure de l'espace paramétrique est précise, considérer la troncature de l'ESB semble

CONCLUSION

être un bon choix.

Pour compléter ces travaux, plusieurs avenues restent à explorer dont une étude de l'admissibilité des estimateurs proposés au Chapitre 2, un sujet qui n'a pas du tout été touché dans ce mémoire. Comme la principale motivation de ces travaux est la recherche d'améliorations de l'ESB, la suite logique est de déterminer si celles-ci sont des estimateurs admissibles. Ce problème pourrait être abordé en établissant des liens entre les estimateurs proposés et les procédures bayésiennes, une tâche qui semble difficile aux premiers abords. De plus, il serait intéressant de déterminer si, moyennant quelques ajustements, un pendant discret de ces travaux peut être développé. Les mélanges de lois de Poisson seraient des applications importantes de cette généralisation. Quelques piste prometteuses ont été explorées à la sous-section 1.4.2 pour les lois de Poisson et binomiale négative, où des résultats originaux ont été présentés pour cette dernière distribution.

Aussi, à la fin du Chapitre 3, on retrouve une brève présentation sur les inférences bayésienne et prédictive pour la loi chi-deux décentrée sous un *a priori* gamma, c'est-à-dire pour un espace paramétrique non borné. Une étude plus approfondie où l'on considère d'autres lois *a priori*, particulièrement en présence d'une contrainte de la forme $\lambda \geq \lambda_0$ avec $\lambda_0 > 0$, serait souhaitable. Toutefois, il est fort probable qu'il soit nécessaire d'utiliser des méthodes numériques telles que les MCMC puisque les lois *a priori* ne seront pas conjuguées. De plus, il serait intéressant de comparer les risques fréquentistes de estimateurs bayésiens, incluant le mode *a posteriori* généralement considéré dans la littérature sur la distribution de Rice, à ceux de l'EVM et de la classe d'estimateurs obtenue au Chapitre 2.

Bibliographie

- [1] D. E. AMOS.
« Computation of Modified Bessel Functions and Their Ratios ».
Mathematics of Computation, 28(125):239–251, 1974.
- [2] James O BERGER, Anne PHILIPPE et Christian P ROBERT.
« Estimation of quadratic functions : noninformative priors for non-centrality parameters ».
Statistica Sinica, 8(2):359–375, 1998.
- [3] J.Calvin BERRY.
« Minimax estimation of a bounded normal mean vector ».
Journal of Multivariate Analysis, 35(1):130–139, 1990.
- [4] Bradley P. CARLIN et Thomas A. LOUIS.
Bayesian methods for data analysis.
Chapman and Hall/CRC, 2008.
- [5] Mo Suk CHOW.
« A complete class theorem for estimating a noncentrality parameter ».
The Annals of Statistics, 15(2):800–804, 1987.
- [6] Arthur P DEMPSTER, Nan M LAIRD et Donald B RUBIN.
« Maximum likelihood from incomplete data via the EM algorithm ».
Journal of the Royal Statistical Society : Series B (Methodological), 39(1):1–22, 1977.
- [7] BS EVERITT et DJ HAND.
« Finite mixture distributions ».
Monographs on Applied Probability and Statistics, London : Chapman and Hall, 1981, 1981.

BIBLIOGRAPHIE

- [8] Thomas S. FERGUSON.
Mathematical Statistics : A Decision Theoretical Approach.
Academic Press, 1967.
- [9] Dominique FOURDRINIER et Éric MARCHAND.
« On Bayes estimators with uniform priors on spheres and their comparative performance with maximum likelihood estimators for estimating bounded multivariate normal means ».
Journal of Multivariate Analysis, 101(6):1300–1399, 2010.
- [10] Ian GOODFELLOW, Yoshua BENGIO et Aaron COURVILLE.
Deep Learning.
MIT Press, 2016.
<http://www.deeplearningbook.org>.
- [11] Ian R. GORDON et Peter HALL.
« Estimating a parameter when it is known that the parameter exceeds a given value ».
Australian & New Zealand Journal of Statistics, 51(4):449–460, 2009.
- [12] Hákon GUDBJARTSSON et Samuel PATZ.
« The Rician distribution of noisy MRI data ».
Magnetic resonance in medicine, 34(6):910–914, 1995.
- [13] Norman L. JOHNSON, Adrienne W. KEMP et Samuel KOTZ.
Univariate Discrete Distributions.
Wiley Series in Probability and Statistics. Wiley, 2005.
- [14] Norman L. JOHNSON, Samuel KOTZ et N. BALAKRISHNAN.
Continuous Multivariate Distributions, Volume 1.
Wiley Series in Probability and Statistics. Wiley, 1994.
- [15] Norman L. JOHNSON, Samuel KOTZ et N. BALAKRISHNAN.
Continuous Multivariate Distributions, Volume 2.
Wiley Series in Probability and Statistics. Wiley, 1995.
- [16] Bobby L JONES, Daniel S NAGIN et Kathryn ROEDER.
« A SAS procedure based on mixture models for estimating developmental tra-

BIBLIOGRAPHIE

- jectories ».
- Sociological methods & research*, 29(3):374–393, 2001.
- [17] DW KIM, C KIM et DH LIM.
« Bayesian approach for Rician non-local means denoising in MR images ».
The Imaging Science Journal, 63(6):303–314, 2015.
- [18] J. F. C. KINGMAN.
Poisson Processes.
Numéro Vol. 3 dans Oxford Studies in Probability. Clarendon Press, 1993.
- [19] Vladimir I KOSTYLEV.
« Energy detection of a signal with random amplitude ».
Dans *2002 IEEE International Conference on Communications. Conference Proceedings. ICC 2002 (Cat. No. 02CH37333)*, volume 3, pages 1606–1610. IEEE, 2002.
- [20] T. KUBOKAWA, C. P. ROBERT et A. K. Md. E. SALEH.
« Estimation of Noncentrality Parameters ».
The Canadian Journal of Statistics / La Revue Canadienne de Statistique, 21(1):45–57, 1993.
- [21] Tatsuya KUBOKAWA.
« A Unified Approach to Improving Equivariant Estimators ».
The Annals of Statistics, 22(1):290–299, 1994.
- [22] Tatsuya KUBOKAWA.
« Shrinkage and modification techniques in estimation of variance and the related problems : A review ».
Communications in Statistics - Theory and Methods, 28(3-4):613–650, 1999.
- [23] Tatsuya KUBOKAWA, Éric MARCHAND et William E. STRAWDERMAN.
« A Unified Approach to Estimation of Noncentral Parameters, the Multiple Correlation Coefficient, and Mixture Models ».
Mathematical Methods of Statistics, 26(2):134–148, 2017.
- [24] Tatsuya KUBOKAWA, Éric MARCHAND, William E. STRAWDERMAN et Jean-Philippe TURCOTTE.
« Minimaxity in predictive density estimation with parametric constraints ».
Journal of Multivariate Analysis, 116:382–397, 2013.

BIBLIOGRAPHIE

- [25] Erich L. LEHMANN et George CASELLA.
Theory of Point Estimation.
Springer-Verlag New York, 1998.
- [26] Erich L. LEHMANN et Joseph P. ROMANO.
Testing Statistical Hypothesis.
Springer-Verlag New York, 2005.
- [27] Qizhai LI, Junjian ZHANG et Shuai DAI.
« On estimating the non-centrality parameter of a chi-squared distribution ».
Statistics & Probability Letters, 79(1):98 – 104, 2009.
- [28] Aziz L'MOUDDEN.
« Sur une approche décisionnelle pour l'analyse bayésienne et l'estimation de densités prédictives ».
Thèse de doctorat, Département de mathématiques, Université de Sherbrooke, Sherbrooke, Québec, Canada, octobre 2018.
<https://savoirs.usherbrooke.ca/handle/11143/14086>.
- [29] Mark MANDELKERN.
« Setting Confidence Intervals for Bounded Parameters ».
Statistical Science, 17(2):149–172, 2002.
- [30] Éric MARCHAND et François PERRON.
« On the minimax estimator of a bounded normal mean ».
Statistics & Probability Letters, 58(4):327–333, July 2002.
- [31] Éric MARCHAND et William E STRAWDERMAN.
« A unified minimax result for restricted parameter spaces ».
Bernoulli, 18(2):635–643, 2012.
- [32] Albert W. MARSHALL, Ingram OLKIN et Barry C. ARNOLD.
Inequalities : Theory of Majorization and Its Application.
Springer-Verlag New York, 2011.
- [33] Itay MAYROSE, Nir FRIEDMAN et Tal PUPKO.
« A Gamma mixture model better accounts for among site rate heterogeneity ».
Bioinformatics, 21(suppl_2):ii151–ii158, 2005.

BIBLIOGRAPHIE

- [34] Geoffrey MCLACHLAN et David PEEL.
Finite mixture models.
John Wiley & Sons, 2004.
- [35] Geoffrey J MCLACHLAN et Kaye E BASFORD.
Mixture models : Inference and applications to clustering.
volume 84. M. Dekker New York, 1988.
- [36] Paul D McNICHOLAS.
Mixture model-based classification.
Chapman and Hall/CRC, 2016.
- [37] Robb J. MUIRHEAD.
Aspects of multivariate statistical theory.
New York : Wiley, 1982.
- [38] Robb J. MUIRHEAD.
« Estimating a Particular Function of the Multiple Correlation Coefficient ».
Journal of the American Statistical Association, 80(392):923–925, 1985.
- [39] Karl PEARSON.
« Contributions to the mathematical theory of evolution ».
Philosophical Transactions of the Royal Society of London. A, 185:71–110, 1894.
- [40] Mathieu PIGEON et Michel DENUIT.
« Composite Lognormal–Pareto model with random threshold ».
Scandinavian Actuarial Journal, 2011(3):177–192, 2011.
- [41] P Wayne POWER et Johann A SCHOONEES.
« Understanding background mixture models for foreground segmentation ».
Dans *Proceedings image and vision computing New Zealand*, volume 2002, pages
10–11, 2002.
- [42] SO RICE.
« Distribution of the duration of fades in radio transmission : Gaussian noise
model ».
BELL system technical journal, 37(3):581–635, 1958.
- [43] Christian ROBERT.
« Modified Bessel functions and their applications in probability and statistics ».
Statistics & Probability Letters, 9(2):155–161, 1990.

BIBLIOGRAPHIE

- [44] T. ROBERTSON, F.T. WRIGHT et R. DYKSTRA.
Order Restricted Statistical Inference.
Probability and Statistics Series. Wiley, 1988.
- [45] Walter RUDIN.
Principles of Mathematical Analysis.
McGraw-Hill Higher Education, 1976.
- [46] K. M. Lal SAXENA et Khursheed ALAM.
« Estimation of the Non-centrality Parameter of a Chi Squared Distribution ».
The Annals of Statistics, 10(3):1012–1016, 1982.
- [47] Peter SCHLATTMANN.
Medical applications of finite mixture models.
Springer, 2009.
- [48] Jan SIJBERS, Arnold J den DEKKER, Johan VAN AUDEKERKE, Marleen VERHOYE et Dirk VAN DYCK.
« Estimation of the noise in magnitude MR images ».
Magnetic resonance imaging, 16(1):87–90, 1998.
- [49] William E. STRAWDERMAN.
« Minimavity ».
Journal of the American Statistical Association, 95(452):1364–1368, 2000.
- [50] Alan Mathison TURING.
« On computable numbers, with an application to the Entscheidungsproblem ».
Proceedings of the London mathematical society, 2(1):230–265, 1937.
- [51] Constance van EEDEN.
Restricted Parameter Space Estimation Problems : Admissibility and Minimavity Properties.
Springer-Verlag New York, 2006.
- [52] Geoffrey S. WATSON.
Statistics on spheres.
New York : Wiley, 1983.
- [53] Éric MARCHAND.
« On moments of beta mixtures, the noncentral beta distribution, and the coeffi-

BIBLIOGRAPHIE

- cient of determination ».
Journal of Statistical Computation and Simulation, 59(2):161–178, 1997.
- [54] Éric MARCHAND.
« Point Estimation of the Coefficient of Determination ».
Statistics and Decisions, 19:137–154, 2001.
- [55] Éric MARCHAND et François PERRON.
« Improving on the MLE of a Bounded Normal Mean ».
The Annals of Statistics, 29(4):1078–1093, 2001.
- [56] Éric MARCHAND et William E. STRAWDERMAN.
« Estimation in Restricted Parameter Spaces : A Review ».
IMS Lecture Notes Monograph, 45:1–24, 2004.
- [57] Éric MARCHAND et William E. STRAWDERMAN.
« Improving on the Minimum Risk Equivariant Estimator of a Location Parameter which is Constrained to an Interval or a Half-interval ».
Annals of the Institute of Statistical Mathematics, 57(1):129–143, 2005.